

Development, Validation, and Interlaboratory Evaluation of a Quantitative Multiplexing Method To Assess Levels of Ten Endogenous Allergens in Soybean Seed and Its Application to Field Trials Spanning Three Growing Seasons

Ryan C. Hill,^{*,†} Trent J. Oman,[†] Xiujuan Wang,[†] Guomin Shan,[†] Barry Schafer,[†] Rod A. Herman,[†] Rowel Tobias,[‡] Jeff Shippar,[§] Bhaskar Malayappan,^{||} Li Sheng,[⊥] Austin Xu,[¶] and Jason Bradshaw[□]

[†]Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268, United States

[‡]EAG Laboratories, 4780 Discovery Drive, Columbia, Missouri 65201, United States

[§]Covance Laboratories, 3301 Kinsman Blvd., Madison, Wisconsin 53704, United States

^{||}Critical Path Services LLC, 3070 McCann Farm Drive, Garnet Valley, Pennsylvania 19060, United States

[⊥]EPL Bioanalytical Services, 9095 W. Harristown Blvd, Niantic, Illinois 62551, United States

[¶]Primerica Analytical Solutions, 259 Wall Street, Princeton, New Jersey 08540, United States

[□]Pyxant Laboratories, 4720 Forge Road #106, Colorado Springs, Colorado 80907, United States

Supporting Information

ABSTRACT: As part of the regulatory approval process in Europe, comparison of endogenous soybean allergen levels between genetically engineered (GE) and non-GE plants has been requested. A quantitative multiplex analytical method using tandem mass spectrometry was developed and validated to measure 10 potential soybean allergens from soybean seed. The analytical method was implemented at six laboratories to demonstrate the robustness of the method and further applied to three soybean field studies across multiple growing seasons (including 21 non-GE soybean varieties) to assess the natural variation of allergen levels. The results show environmental factors contribute more than genetic factors to the large variation in allergen abundance (2- to 50-fold between environmental replicates) as well as a large contribution of Gly m 5 and Gly m 6 to the total allergen profile, calling into question the scientific rationale for measurement of endogenous allergen levels between GE and non-GE varieties in the safety assessment.

KEYWORDS: surrogate peptide, multiplexing method, method validation, endogenous allergens, soybean, genetically engineered, LC–MS/MS, mass spectrometry, food safety, Glycine max

INTRODUCTION

Soybean (*Glycine max*) is a crop rich in protein and oil, making it a very profitable commodity.¹ However, soybean has been recognized as a source of IgE-mediated food allergies² and can be found in food products ranging from margarine to cooking oils. To facilitate increasing global demand, genetically engineered (GE) crops containing beneficial traits to protect plants from insect and disease damage as well as to provide tolerance to herbicides have been developed.^{3–5} GE crops require regulatory approval before they can be sold to the general public. As part of the regulatory approval process for the food and feed safety assessment, comprehensive compositional assessments are performed to compare GE crops with their conventional counterparts to determine if meaningful differences in key nutrients, metabolites, or known toxins and allergens exist.^{6,7}

The European Commission has mandated (European Commission Implementing Regulation 503/2013) that potential soybean endogenous allergens be included in the compositional safety assessment.^{8–12} Of the 15 potential allergens listed, sufficient data supporting classification as an allergen and/or characterized protein sequence information is lacking for Gly m 2 (defensin), lectin, lipoxygenase, 50 kDa protein, 39 kDa protein,

and P-22–25.¹³ However, for the ten remaining soybean allergens (Gly m 1, hydrophobic seedpod protein; Gly m 3, profilin; Gly m 4, SAM22/pathogenesis related protein; Gly m 5, β -conglycinin/storage protein; Gly m 6, glycinin/storage protein; Gly m 8, 2S albumin/prolamin; Gly m Bd 28 K, Asn-linked glycoprotein; Gly m Bd 30 K, P34/protease; Kunitz trypsin inhibitors 1 and 3), literature exists to support possible food allergenicity,¹⁴ and quantitative methods may need to be developed to support the safety assessment.

Technical challenges for quantitative measurement of endogenous soybean allergens are well-documented.¹⁵ Historically, uniplex immunochemistry methods such as enzyme-linked immunosorbent assay (ELISA)¹⁶ have been used to quantify proteins in complex matrixes; however, the methods are costly and challenging to develop for endogenous analytes in terms of antibody specificity to the target allergen (sequence homology), selectivity (nontarget endogenous components not suitable for

Received: March 5, 2017

Revised: June 12, 2017

Accepted: June 20, 2017

Published: June 21, 2017



Table 1. Specific Instrumental Conditions for Endogenous Soybean Allergens

peptides monitored	charge state	precursor ion(<i>m/z</i>)	product ion (<i>m/z</i>)	DP (V)	CE (V)
Gly m 1, SYPSNATCPR, quant	+2/ <i>y</i> 8	548.25	845.39	71	29
Gly m 1, SYPSNATCPR, conf	+2/ <i>y</i> 7	548.25	748.34	71	29
Gly m 1, SYPSNATCPR, heavy	+2/ <i>y</i> 8	553.25	855.40	71	29
Gly m 3, YMVIQGEPEGAVIR, quant	+2/ <i>y</i> 8	716.88	798.45	83	35
Gly m 3, YMVIQGEPEGAVIR, conf	+2/ <i>y</i> 9	716.88	926.51	83	35
Gly m 3, YMVIQGEPEGAVIR, heavy	+2/ <i>y</i> 8	721.89	808.46	83	35
Gly m 4, ALVTDADNVIPK, quant	+2/ <i>y</i> 9	628.35	972.50	77	32
Gly m 4, ALVTDADNVIPK, conf	+2/ <i>y</i> 8	628.35	871.45	77	32
Gly m 4, ALVTDADNVIPK, heavy	+2/ <i>y</i> 9	632.36	980.51	77	32
Gly m 5, NILEASYDTK, quant	+2/ <i>y</i> 8	577.29	926.45	73	30
Gly m 5, NILEASYDTK, conf	+2/ <i>y</i> 7	577.29	813.36	73	30
Gly m 5, NILEASYDTK, heavy	+2/ <i>y</i> 8	581.30	934.46	73	30
glycinin G1, VLIVPQNFWAAR, quant	+2/ <i>y</i> 9	713.43	1001.55	83	35
glycinin G1, VLIVPQNFWAAR, conf	+2/ <i>y</i> 10	713.43	1100.62	33	35
glycinin G1, VLIVPQNFWAAR, heavy	+2/ <i>y</i> 9	718.44	1011.56	83	35
glycinin G2, VTAPAMR, quant	+2/ <i>y</i> 5	373.21	545.29	58	22
glycinin G2, VTAPAMR, conf	+2/ <i>y</i> 4	373.21	474.25	58	22
glycinin G2, VTAPAMR, heavy	+2/ <i>y</i> 5	378.21	555.30	58	22
glycinin G3, NNNPFSFLVPPK, quant	+2/ <i>y</i> 9	687.36	1031.59	81	34
glycinin G3, NNNPFSFLVPPK, conf	+2/ <i>y</i> 10	687.36	1145.64	81	34
glycinin G3, NNNPFSFLVPPK, heavy	+2/ <i>y</i> 9	691.37	1039.61	81	34
glycinin G4, VESEGGLIQTWNSQHPELK, quant	+3/ <i>y</i> 10	718.03	1239.61	84	37
glycinin G4, VESEGGLIQTWNSQHPELK, conf	+3/ <i>y</i> 9	718.03	1138.56	84	37
glycinin G4, VESEGGLIQTWNSQHPELK, heavy	+3/ <i>y</i> 10	720.70	1247.63	84	37
glycinin precursor, NGLHLPSYSPYR, quant	+2/ <i>y</i> 8	750.88	966.47	86	36
glycinin precursor, NGLHLPSYSPYR, conf	+2/ <i>y</i> 9	750.88	1079.55	36	36
glycinin precursor, NGLHLPSYSPYR, heavy	+2/ <i>y</i> 8	755.89	976.48	86	36
Kunitz trypsin inhibitor 1, GGGIEVDSTGK, quant	+2/ <i>y</i> 7	510.25	735.35	68	27
Kunitz trypsin inhibitor 1, GGGIEVDSTGK, conf	+2/ <i>y</i> 6	510.25	606.31	68	27
Kunitz trypsin inhibitor 1, GGGIEVDSTGK, heavy	+2/ <i>y</i> 7	514.26	743.37	68	27
Kunitz trypsin inhibitor 3, GIGTIISPYR, quant	+2/ <i>y</i> 6	582.33	722.38	74	30
Kunitz trypsin inhibitor 3, GIGTIISPYR, conf	+2/ <i>y</i> 9	582.33	993.54	74	30
Kunitz trypsin inhibitor 3, GIGTIISPYR, heavy	+2/ <i>y</i> 6	587.33	732.39	74	30
Gly m Bd 28 K, NKPQFLAGAASLLR, quant	+3/ <i>y</i> 7	495.96	687.42	67	25
Gly m Bd 28 K, NKPQFLAGAASLLR, conf	+3/ <i>y</i> 8	495.96	758.45	67	25
Gly m Bd 28 K, NKPQFLAGAASLLR, heavy	+3/ <i>y</i> 7	499.29	697.42	67	25
Gly m Bd 30 K, GVITQVK, quant	+2/ <i>y</i> 5	372.73	588.37	58	22
Gly m Bd 30 K, GVITQVK, conf	+2/ <i>y</i> 4	372.73	475.29	58	22
Gly m Bd 30 K, GVITQVK, heavy	+2/ <i>y</i> 5	376.74	596.39	58	22
Gly m 8, IMENQSEEELEEK, quant	+2/ <i>y</i> 7	739.84	863.40	85	36
Gly m 8, IMENQSEEELEEK, conf	+2/ <i>y</i> 9	739.84	1105.50	85	36
Gly m 8, IMENQSEEELEEK, heavy	+2/ <i>y</i> 7	743.85	871.41	85	36

proteins with different isoforms), limited quantitative range (allergens are typically highly abundant), and required purification of protein standards. A two-dimensional liquid chromatography with ultraviolet and mass spectrometry detection (2DLC–UV/MS)¹⁷ method was developed to detect multiple isoforms for Gly m 4 from soybean seed; however, the methodology is technically complex and time-consuming to perform, has low sensitivity, and also requires protein standard purification. Additionally, there are few studies demonstrating the expression range of endogenous allergen levels in conventional non-GE soybean for a useful comparison.¹⁸ Thus, a simple and robust multiplexing assay is beneficial for generating consistent data over time to aid risk assessors in interpreting results. A multiplex surrogate peptide assay using liquid chromatography with tandem mass spectrometry (LC–MS/MS) was reported by Houston et al.¹⁹ to monitor five allergens (five subunits of Gly m 6, Gly m 5, two trypsin inhibitors, and Gly

m Bd 28 K) and applied to various soybean varieties to demonstrate the utility of the methodology.

In this study, we further applied the LC–MS/MS technology and developed a multiplexing method to measure ten soybean allergens, including Gly m 1, Gly m 3, Gly m 4, Gly m 5, Gly m 6 (glycinin G1, glycinin G2, glycinin G3, glycinin G4, and glycinin precursor individual subunits), Gly m Bd 28 K, Gly m Bd 30 K, Kunitz trypsin inhibitor 1, Kunitz trypsin inhibitor 3, and Gly m 8 across as many identified isoforms as possible. Limited guidance exists for the validation of LC–MS methods for protein quantitation,^{20,21} especially for endogenous analytes. The implemented validation protocol for endogenous soybean allergens is discussed. In addition, the analytical method was transferred to six laboratories for intralaboratory and interlaboratory assessment of the analytical method. This method was further applied to 3 soybean field studies harvested from multiple locations between 2011 and 2015 which included 21

commercially available non-GE soybean varieties from which the natural variation of endogenous soybean allergen levels were assessed.

MATERIALS AND METHODS

Materials. Synthetic peptides representative of natural abundance peptides from each soybean allergen and corresponding heavy labeled peptides (L-lysine- $^{13}\text{C}_6$, $^{15}\text{N}_2$ or L-arginine- $^{13}\text{C}_6$, $^{15}\text{N}_4$) were obtained from New England Peptide (Gardner, MA). Sequencing grade trypsin was purchased frozen from Promega Corporation (Madison, WI).

Methodology. The methodology can be applied to a single sample or large field trials. Soybean seed(s) were ground, lyophilized, weighed into 100 (± 0.5) mg aliquots and stored at -80°C until analysis. Each aliquot was defatted with 1 mL of hexane, vortex mixed for 1 min, and centrifuged for 2 min at >3000 rpm and 4°C . The supernatant was discarded, and the defatting procedure was repeated followed by evaporation of residual hexane in a fume hood for 1 h. The resulting sample was extracted with 1 mL of extraction buffer (5 M urea, 2 M thiourea, 50 mM Tris-HCl, and 65 mM dithiothreitol) for 1 h in a thermomixer set at 1100 rpm at ambient temperature. The sample was centrifuged for 10 min at >3000 rpm and 4°C . The supernatant was collected and diluted 10 $\times$ with HPLC water to bring allergen concentrations into the peptide calibration range. A 10 μL aliquot of diluted extract was denatured at 95°C for 20 min and reduced with the addition of 4 μL of 0.5 M dithiothreitol followed by refrigeration at 4°C for 10 min. The denatured extract was further diluted 15 $\times$ with HPLC water and pH buffered with 20 μL of 1 M Tris-HCl followed by overnight incubation (~ 15 h) at 37°C with 5 μg of trypsin enzyme. Following digestion, a multistock of heavy isotope labeled peptide internal standards was added to the digested extract (to normalize samples) at a final concentration of 250 ng/mL. The digestion reaction was quenched with 10 μL formic acid/water (50/50, v/v) and centrifuged for 10 min at >3000 rpm and 4°C . An aliquot of digested extract was analyzed along with calibration standards containing both synthetic natural abundance peptides and heavy isotope labeled peptide internal standards in buffer (HPLC water containing 0.1% bovine serum albumin) by LC–MS/MS.

Instrumental Conditions. The LC–MS/MS system included a Waters Acquity I-Class UPLC (Milford, MA) coupled to AB Sciex 6500 QTRAP (Framingham, MA) with turbo ion-spray source. The autosampler temperature was kept at 4°C during analysis. The reverse phase analysis employed an Acquity UPLC BEH130 C18 1.7 μm (2.1×50 mm) column held at 50°C . The organic mobile phase (MPB) was acetonitrile containing 0.1% formic acid, and the aqueous mobile phase (MPA) was water containing 0.1% formic acid. The LC flow rate was 0.4 mL/min with a linear gradient from 2 to 28% MPB over 10 min, followed by 98% MPB for 1 min to wash the column and 2% MPB for 1 min to re-equilibrate the column. The needle wash containing methanol/water (50/50, v/v) was run for 1 min with no detected carryover between samples. The mass spectrometer was operated in positive ion mode with ionspray voltage of 5500 V and temperature set to 450°C . The MS/MS transitions for each peptide can be found in Table 1 with respective optimized collision and depolarization energies obtained from skyline proteomics software. Other relevant instrument parameters included curtain gas of 35 psi, ion source gas 1 and 2 of 55 psi, collision gas set to medium, entrance potential of 10 V, and collision cell exit potential of 15 V.

Method Validation Parameters. The analytical method validation was adapted to endogenous analytes from techniques performed similarly for transgenic proteins.²² The following validation experiments were performed with at least three commercially available soybean varieties to incorporate different genomic backgrounds:

- The extraction efficiency of endogenous soybean allergens from soybean seed was evaluated by repeatedly extracting the soybean seed tissue samples five times and determining the amount of endogenous soybean allergen protein from each extraction following the analytical method. The extraction efficiency was calculated as a percentage of the amount detected in the initial extraction divided by the sum from all extractions.

- The completeness of digestion was examined with a time course experiment. Soybean seed was extracted following the analytical procedure and digested for 4, 8, 10, 12, 15, 18, 20, 24, and 28 h time increments in triplicate. Each digest time point was analyzed for endogenous soybean allergen concentrations by LC–MS/MS. An SDS-PAGE gel was performed to confirm completeness of digestion for the 0, 4, 8, 10, 12, and 15 h time points.
- Method accuracy, precision, and ruggedness were assessed by determining the recovery of synthetic natural abundance peptides representing each endogenous soybean allergen that had been fortified into digested soybean seed over three separate experiments (assessment included three different soybean varieties within each experiment). Soybean seed was extracted and digested following the digestion protocol. Digested soybean tissue was pooled together, and synthetic natural abundance peptides were fortified at multiple concentration levels, which included the target limit of quantitation (LOQ) for each peptide and the middle and high concentration levels of the calibration curve. For each analytical set, three aliquots from the pooled digest were analyzed to determine the mean baseline response for recovery due to endogenous presence of analytes. The amount of surrogate peptide representing each soybean allergen was interpolated from the peptide calibration curve. The assay accuracy was indicated as percent of recovery with acceptable recovery between 70 and 120% and RSD $< 20\%$ using the following equation: $\text{recovery (\%)} = \frac{\text{gross detected peptide} \times [\text{fortified peptide level} / (\text{fortified peptide level} + \text{detected mean baseline})]}{\text{gross detected peptide}} \times 100$.
- The possible matrix effects were tested by comparing response (peak area) of heavy isotope peptide spiked into buffer (deionized water containing 0.1% bovine serum albumin) compared with response of heavy isotope peptide spiked into digested soybean seed matrix. In general, a difference of $>20\%$ in the observed response between the heavy isotope peptide in buffer and digested soybean seed could be considered indicative of a potential matrix effect. Matrix effect was calculated as a percentage using the following equation: $\text{matrix effect (\%)} = \frac{[(\text{average peak area of heavy isotope peptide spike in matrix} / \text{average peak area of heavy isotope peptide spike in buffer}) - 1] \times 100}{1}$.
- A specificity examination of each surrogate peptide to respective endogenous soybean allergen was performed by extracting soybean seed varieties and analyzing them by LC–MS/MS. The peak areas of the quantitative peptide monitored in the method and MS/MS transitions from at least one more peptide identified during method development were compared.
- Dilution linearity of endogenous soybean allergens was evaluated using serial dilutions (1–32 $\times$ dilution) of extracted soybean supernatant. The amount of endogenous soybean allergen was interpolated from the quantitative range of the standard curve. A relative standard deviation less than 20% demonstrated protein dilution linearity for each soybean allergen.
- Digested peptide stability was examined with digested soybean seed fortified with synthetic natural abundance peptides along with calibration standards analyzed on day 0 and stored at 4°C (temperature of the autosampler in the analytical method parameters). After 3 days, the samples were reanalyzed using the same instrument parameters as the day 0 analysis with freshly prepared standards and held to the same criteria as the recovery experiments.

Ring Trial. An independent assessment of the LC–MS/MS method was performed at six participating laboratories: EAG Laboratories (Columbia, MO), Covance Laboratories (Madison, WI), Critical Path Services (Garnet Valley, PA), EPL Bioanalytical Services (Niantic, IL), Primera Analytical Solutions (Princeton, NJ), and Pyxant Laboratories (Colorado Springs, CO). Seven preweighed 100 mg aliquots from one soybean sample, synthetic natural abundance peptides, and heavy isotope labeled peptides were provided to each of the laboratories. Three soybean aliquots were used for preliminary testing of the method,



Figure 1. Example generation of Gly m 1 consensus protein sequence used for surrogate peptide selection to cover all possible isoforms (applied similarly for each endogenous soybean allergen). Pework identified all known sequences for Gly m 1 from public databases (UniProt, NCBI, and Phytozome). Sequence alignments were performed using Vector NTI Align X software, and overall sequence homology between isoforms was assessed. The nonsimilar, conservative, identical, and weakly similar residues are highlighted in white, blue, yellow, and green, respectively. The Gly m 1 peptide SYPSNATCPR was chosen as a surrogate with the greatest isoform coverage. The SYPSNATCPR was then validated to quantitatively measure Gly m 1 with appropriate sensitivity, selectivity, precision, and accuracy.

Table 2. Soybean Allergen Identity, Consensus Sequences, and Peptide Information^a

allergen nomenclature	IgE-binding protein	accession number	peptide monitored	protein molecular weight (Da)	peptide molecular weight (Da)
Gly m 1	hydrophobic proteins	ABA54898.1	SYPSNATCPR	12 482.64	1095.20
Gly m 3	profilin	CAA11755.1	YMVIQGEPAVIR	14 100.07	1432.70
Gly m 4	SAM22	P26987	ALVTDADNVIK	16 771.81	1255.43
Gly m 5	β -conglycinin	I21281	NILEASYDTK	70 293.13	1153.25
Gly m 6	glycinin (G1 subunit)	P04776	VLIVPQNFVVAAR	55 706.34	1425.74
	glycinin (G2 subunit)	P04405	VTAPAMR	54 390.76	744.91
	glycinin (G3 subunit)	P11828	NNNPFSLVPPK	54 241.73	1373.57
	glycinin (G4 subunit)	P02858	VESEGGLIQTWNSQHPELK	63 587.16	2152.35
	glycinin (glycinin precursor subunit)	Q43452	NGLHLPSPYPR	63 876.47	1500.68
	Kunitz trypsin inhibitor 1	P25272	GGGIEVDSTGK	22 545.94	1019.08
	Kunitz trypsin inhibitor 3	P01070	GIGTHISSPYR	24 005.29	1163.34
Gly m 8	Gly m Bd 28 K	12697782	NKPQFLAGAASLLR	52 944.36	1485.75
	Gly m Bd 30 K	84371705	GVITQVK	42 757.81	743.90
	2S albumin	NP 001238443	IMENQSEEELEK	18 459.97	1478.59

^aProtein and peptide molecular weight (monoisotopic) used to convert peptide into protein results = detected peptide $\times$ (protein MW/peptide MW).

and the remaining four soybean aliquots were used for the formal interlaboratory/intralaboratory method assessment. One sample was dedicated to demonstrating spike recovery following the protocol explained in the method validation section, and the remaining three samples were treated as unknowns and compared across the laboratories. The same methodology and LC–MS/MS parameters developed at Dow AgroSciences were employed by all laboratories regardless of instrumentation platform to reduce variability and demonstrate robustness of the method. Four of the laboratories used

AB Sciex 6500 QTRAP mass spectrometers; one laboratory used an AB Sciex 4000 mass spectrometer, and one laboratory used an Agilent 6460 mass spectrometer.

Field Trials. Three field studies containing 21 commercially available non-GE soybean varieties were harvested in 2011, 2012, and 2015 at 9 or 10 field sites depending on the study (Iowa, 2 sites; Illinois, 2 sites; Indiana and Missouri, 2 sites; Nebraska, 2 sites; and Pennsylvania). At each field site, four replicate plots were planted with plots arranged in a randomized complete block design. Commercially available soybean

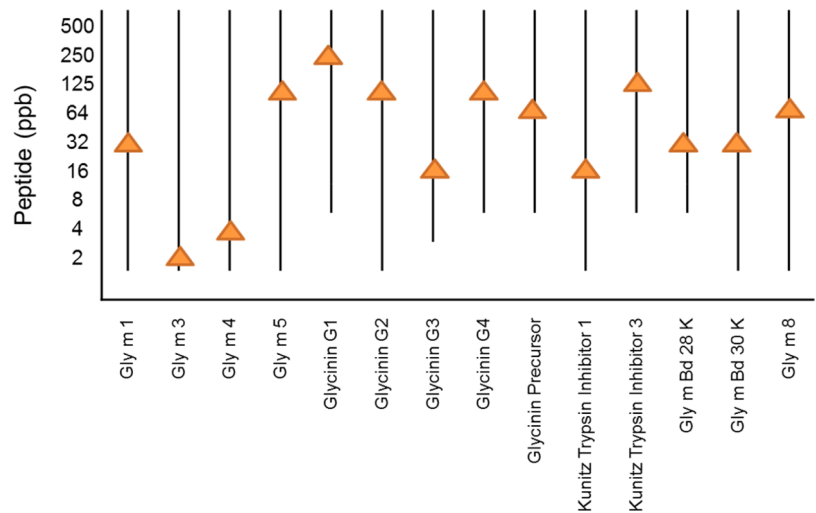


Figure 2. In a single-plex assay, development will result in an optimized linear range for the target analyte. For a multiplexing assay, the difference in sensitivity (LOQ) between all surrogate peptides and the general expression range for each allergen (ranging from ppb to ppm) had to include an appropriate dilution scheme during method development. The peptide sensitivity (vertical line in ppb) compared to typical expression range (indicated as a triangle) is shown above. An 8–10× dilution was found to be optimal to bring all allergens into the multiplex calibration range for all samples analyzed.

Table 3. Endogenous Soybean Allergen Limits of Detection and Quantitation

protein	proposed LOQ recovery ^a (ng/mg)	standard deviation (s)	target LOD (ng/mg)	calculated limit of detection (3s)	target LOQ (ng/mg)	calculated limit of quantitation (10s)
Gly m 1	0.46	0.03	0.23	0.09	0.46	0.30
Gly m 3	0.16	0.03	0.20	0.09	0.39	0.30
Gly m 4	0.42	0.03	0.27	0.09	0.53	0.30
Gly m 5	2.44	0.08	1.22	0.24	2.44	0.80
glycinin G1	14.01	0.65	6.25	1.95	12.50	6.50
glycinin G2	2.84	0.10	1.46	0.30	2.92	1.00
glycinin G3	1.66	0.19	1.58	0.57	3.16	1.90
glycinin G4	5.60	0.76	2.36	2.28	4.73	7.60
glycinin precursor	6.43	0.63	3.41	1.89	6.81	6.30
Kunitz trypsin inhibitor 1	0.85	0.03	0.44	0.09	0.88	0.30
Kunitz trypsin inhibitor 3	4.07	0.79	1.65	2.37	3.30	7.90
Gly m Bd 28 K	4.56	0.40	5.70	1.20	11.40	4.00
Gly m Bd 30 K	2.27	0.12	1.15	0.36	2.30	1.20
Gly m 8	0.50	0.02	0.25	0.06	0.50	0.20

^aAverage recovery = protein concentrations converted from the proposed limit of quantitation natural abundance peptide fortifications in nine samples over three separate analyses.

varieties were randomized across sites with three varieties at each site per study and grown to the R8 (maturity) growth stage for seed harvesting. The non-GE soybean varieties included AgVenture AV 39A0, Beck's 389N, Beck's 401, DSR 3510, DSR 36Y14Y1, DSR 99915, Dyna-Gro 3410SCN, Dyna-Gro V388SCN, HiSOY 38C60, Hoffman H387, IL 3503, LG Seeds C3884N, L&M 34, Mark C1438SB, Pfister 39C74, Pioneer 93Y41, Stine 3822-2, Stine 3900-2, Stine 3920-2, and Williams 82. Approximately 500 g were collected per sample, and 4 replicate samples were obtained per soybean variety at each site. Samples were stored at −80 °C until analysis. Analysis of endogenous soybean allergen content was performed in May 2014 through May 2016 following the validated analytical procedure.

Variance Component Analysis for Endogenous Soybean Allergens. Variance component analysis was performed to assess the contributions of genetic and environmental factors to soybean allergen levels evaluated in the three soybean field trials similar to previously reported compositional analysis studies.^{18,23–25} To estimate the variance component from different sources, soybean allergen levels were analyzed with a linear regression model:

$$Y = \mu + G + E + G^*E + \varepsilon$$

where Y is the observed allergen level, μ is the overall mean, G is the genetic effect from soybean variety, E is the environmental effect, G^*E is the interaction effect of soybean variety and environment, and ε is the error. All of the three factors (G , E , and G^*E) were considered as random effects. The variance components were calculated using the SAS procedure PROC MIXED (SAS 9.3, SAS Institute, Cary, NC) using estimation method REML. The proportion of variance for genetic effect and environmental effect were calculated by dividing the variance component of genetic factor (G) and environmental factor (E) by total variance, respectively. Here, a field location at a given harvest year is considered as an environment.

RESULTS

Method Development. The purpose for the analytical method was to compare allergen levels from novel GE soybean with those of conventional non-GE soybean varieties. Several considerations such as choice of calibration standard, selectivity

Table 4. Extraction Efficiency of Soybean Allergens across Multiple Soybean Varieties^a

protein	soybean variety 1 mean EE (RSD) (%)	soybean variety 2 mean EE (RSD) (%)	soybean variety 3 mean EE (RSD) (%)	overall mean EE (%)
Gly m 1	88 (0.76)	80 (2.85)	86 (0.93)	85
Gly m 3	75 (0.72)	77 (5.66)	83 (8.21)	78
Gly m 4	79 (4.77)	82 (1.06)	85 (2.12)	82
Gly m 5	83 (10.3)	83 (0.57)	86 (6.69)	84
glycinin G1	80 (2.10)	80 (0.70)	82 (1.28)	81
glycinin G2	83 (0.00)	84 (4.17)	87(1.34)	85
glycinin G3	78 (1.45)	80 (4.33)	82 (0.00)	80
glycinin G4	79 (6.02)	81 (6.38)	82(19.1)	80
glycinin precursor	80 (2.14)	79 (6.08)	82 (3.15)	80
Kunitz trypsin inhibitor 1	79(11.0)	84 (0.21)	87 (2.06)	83
Kunitz trypsin inhibitor 3	81 (6.87)	83 (5.62)	84 (2.19)	83
Gly m Bd 28 K	60 (0.70)	64 (2.86)	64 (3.17)	63
Gly m Bd 30 K	80 (2.83)	82 (0.98)	85 (2.94)	82
Gly m 8	81 (2.45)	83 (6.73)	83 (0.68)	83

^aEE = extraction efficiency from three biological replicates serially extracted five times.

of the assay for known isoforms, and extraction conditions broad enough to incorporate a range of protein chemistries for a multiplexing assay were key challenges during assay development. Synthetic natural abundance peptides were chosen as reference standards for the analytical method, as synthetic peptides are easily obtainable and can be readily reproducible across suppliers. Alternatively, protein standards were considered; however, purification challenges for all allergens and the necessity for further characterization for each of standard could lead to variability between laboratories. Heavy isotope labeled peptides were chosen as internal standards to normalize results and were included in the calibration curve and unknown samples. The value of internal standards to reduce analytical variability was reported by Houston et al. using bovine serum albumin as the normalization source. The use of structurally identical heavy isotopes in this assay facilitates peptide specific normalization for accurate interpretation of results across experiments and laboratories.

Peptide Selection. All recognized protein sequences for each allergen were obtained from public databases UniProt, NCBI, and Phytozome. Sequence homology was assessed with Vector NTI Align X software using a CLUSTAL W type alignment to identify shared sequence homology between the allergen isoform sequences. From this analysis, a consensus sequence was determined and used for peptide selection to guarantee the greatest isoform coverage (Figure 1). The consensus sequence was digested with trypsin in silico to generate theoretical peptide fragments to be measured by LC–MS/MS. All possible tryptic peptides from the consensus sequence were evaluated in 13 different soybean varieties (genetic backgrounds) for peptide selection. Peptides were selected primarily on the degree of conservation among the available protein sequences. Further selection criteria included a length of at least 6–20 amino acids to provide high confidence in specific identification and consistency of response from an

overnight digestion. The possibility for missed cleavages was qualitatively evaluated (up to three missed cleavages) to assess the consistency of digestion prior to ordering peptide standards. One peptide was selected per endogenous soybean allergen to use for quantitation to preserve the duty cycle due to the volume of analytes in the analysis. Additionally, two product ions were monitored per peptide for real time confirmation of results. The full list of consensus sequences, peptides, and allergen information can be found in Table 2.

Sample Treatment. Soybean seed was ground using liquid nitrogen and a blender to a consistent fine powder. The consistency of particles in the ground soybean seed was found to improve the precision of results. A hexane defatting procedure was necessary to remove lipids from soybean seed which otherwise affected chromatography and reduced the sensitivity in addition to causing an accumulation of residue on the analytical column over time. A 1 mL aliquot of hexane to submerge the soybean seed followed by a 1 min vortex and centrifugation defatting procedure performed in duplicate was found to be sufficient to remove the lipids. The hexane used in the defatting step was completely removed before extraction to prevent formation of a partitioning organic bilayer found to affect precision of results.

Extraction. A wide range of extraction conditions were evaluated from work reported by Natarajan et al.²⁶ (urea, thiourea/urea, phenol, and modified trichloroacetic acid/acetone) in addition to buffer/seed volume and mechanical grinding versus agitation for adaptation to a multiplexing LC–MS/MS assay. The thiourea/urea extraction (5 M urea, 2 M thiourea, 50 mM Tris pH 8.0, and 65 mM DTT) was found to be optimal for reproducible extraction of the large range of hydrophobicity and abundances observed for all endogenous allergens using a 1 h agitation with sonication in a water bath. The extraction was later optimized to a thermomixer which allowed for a more consistent extraction in a larger throughput format. Filtration of extraction solution after preparation was found to improve consistency of extraction.

Digestion. The high abundance of allergens allowed for reproducible response from an overnight digestion to be evaluated while varying the denature time, enzyme concentration, digestion time, and digestion buffer during development. An incubation of 20 min at 95 °C for denaturation and reduction, enzyme concentration of 5 µg, and overnight incubation for 15 h were found to be sufficient with no increase in peptide response while maintaining peptide stability for all allergens over the course of the digestion.

Sensitivity. The general expression range of endogenous allergens had to be well-characterized during method development to account for the large range of abundance among allergens (ppb to mid-ppm) in addition to sensitivity differences between surrogate peptides (Figure 2). Ideally, the calibration curve would be prepared in control matrix to mimic the environment of the unknown samples. Due to the presence of allergen proteins in the matrix, the calibration curve was prepared in aqueous buffer (HPLC water containing 0.1% bovine serum albumin) with peptide concentrations ranging from 2–500 ppb. The limits of quantitation (Table 3) incorporate the linear response from each peptide and may be optimized lower depending on instrumentation linearity limitations.

Method Validation. Due to the endogenous nature of soybean allergens (absence of control tissue) and the complexity of generating purified protein standards, the validation was performed with sequentially independent experiments (extrac-

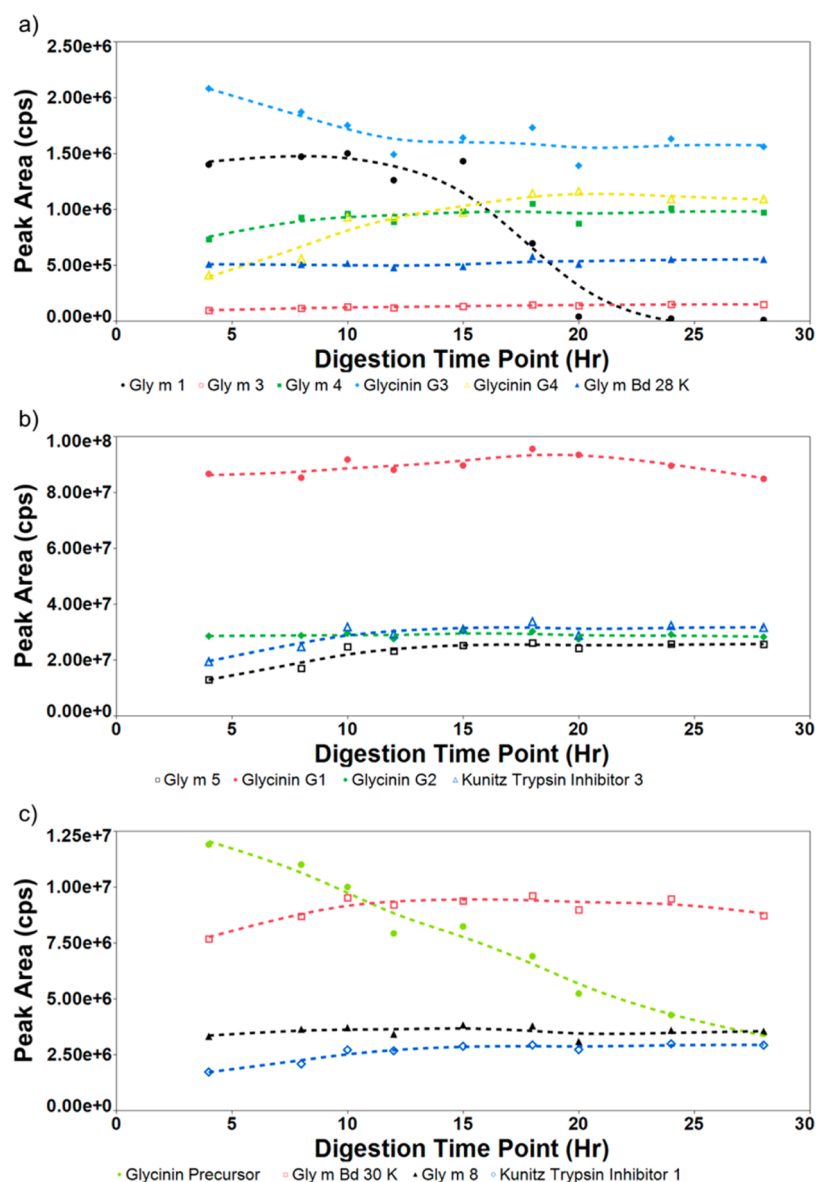


Figure 3. A time course experiment was used to evaluate completeness of digestion with the reaction quenched at 4, 8, 10, 12, 15, 18, 20, 24, 28 h time points in triplicate (x -axis). For a multiplex assay, the digestion protocol will favor the least stable peptide. The digestion plots show mean surrogate peptide response (y -axis) is stable for all peptides except for Gly m 1 and glycinin precursor, which are less stable past the 15 h digestion interval.

tion through detection) with synthetic natural abundance and heavy labeled isotope peptide internal standards which are employed in the calibration standards for the analytical method. For transgenic proteins, the availability of control tissue absent of target analyte allows for a more straightforward examination to validate several method components in a single experiment.

The extraction efficiency examination evaluated three different soybean varieties, each serially extracted five times. The extraction was found to be effective with mean extraction efficiencies ranging from 78 to 85% (except Gly m Bd 28 K at 63%) and consistent with RSD < 20% for all soybean allergens (Table 4). If antibodies had been available, the completeness of extraction could have been evaluated with Western blots on the exhaustively extracted pellet. The validated extraction procedure across 14 (including the 5 subunits for glycinin) diverse protein chemistries demonstrates the potential to add additional allergens in the future with minimal method development.

Although the measured peak area ratio of natural abundance peptide to heavy isotope internal standard should account for matrix effects and other instrumentation fluctuations on the sample, the potential matrix effects were examined. The endogenous soybean allergen proteins are present in the matrix; hence, the effect of matrix on the natural abundance peptides representing each endogenous soybean allergen could not be directly assessed. Fortunately, the heavy labeled isotope peptide internal standards are not present in matrix and coelute with natural abundance peptides, allowing for an estimation of the potential matrix effect on the natural abundance peptides in soybean seed matrix. All matrix effects were determined to be below 20%, indicating minimal effect on quantitative interpretation of the natural abundance peptide.

A well-characterized digestion protocol is critical for accurate surrogate peptide quantitation. Ideally, evaluation would be performed using control tissue of a similar background to the test substance and purified protein standards to validate a digestion

Table 5. Dilution Linearity Evaluation of Endogenous Soybean Allergens

protein	dilutions ^a	mean (μg/mg)	std dev	RSD (%)	range (μg/mg)
Gly m 1	5	0.41	0.03	8	0.37–0.48
Gly m 3	5	0.05	0.01	14	0.04–0.06
Gly m 4	5	0.29	0.03	11	0.24–0.35
Gly m 5	5	66.07	7.67	12	55.30–81.92
glycinin G1	5	14.65	2.09	14	11.08–17.66
glycinin G2	5	31.52	1.94	6	28.77–35.27
glycinin G3	5	3.34	0.43	13	2.78–4.47
glycinin G4	5	13.67	2.18	16	9.97–17.57
glycinin precursor	5	4.04	0.69	17	2.67–5.04
Kunitz trypsin inhibitor 1	5	0.87	0.07	8	0.73–1.03
Kunitz trypsin inhibitor 3	5	8.95	0.77	9	7.72–10.56
Gly m Bd 28 K	5	0.78	0.09	11	0.66–1.41
Gly m Bd 30 K	5	6.17	0.47	8	5.27–6.89
Gly m 8	5	6.73	0.62	9	5.40–8.16

^aExtract dilutions spanning 1–32×.

protocol. Due to absence of control tissue and the challenges associated with purifying protein standards, a time course experiment was used to evaluate completeness of digestion.^{27,28} Soybean grain was extracted and digested, and the reaction was quenched at 4, 8, 10, 12, 15, 18, 20, 24, and 28 h time points in triplicate. In a multiplex assay, the digestion protocol will tend to favor the least stable peptide. The digestion plots in Figure 3 show surrogate peptide response is stable for all peptides except for Gly m 1 and glycinin precursor, which are less stable past the 15 h digestion interval. The digestion time points were confirmed by SDS-PAGE and show a high abundance of higher molecular weights at time 0 which are absent in the 15 h time point (not shown). For the 10 soybean allergens (and all 5 subunits representing Gly m 6), the 15 h digestion protocol was determined to be acceptable with stable peptide response at the 15 h time point.

Dilution linearity of extracted target proteins was assessed spanning 1–32× dilution of extract supernatant to validate

practical dilutions applied to the large range of allergen abundances. In addition, a 21× method factor introduced during the sample preparation translates to a 21–735× dilution at the peptide level. The dilution linearity for endogenous allergens was acceptable with RSD ≤ 20% (Table 5). Considering a multiplex analysis, a dilution range of 8–12× was found to be optimal to maintain the most abundant proteins (Gly m 5 and Gly m 6) and least abundant allergens (Gly m 3) in the calibration range.

The method is selective for the determination of endogenous soybean allergens by chromatographic separation coupled to MS/MS detection. When detection is by tandem mass spectrometry, confirmation of the presence of the analyte should require the observation of a precursor ion plus two structurally significant product ions observed at the same retention time.²⁹ In addition to monitoring one precursor ion and two product ions for each peptide, corresponding heavy isotope labeled peptide internal standards were used and observed at the same retention time as the monitored natural abundance peptides to normalize all samples and calibration standards. The method selectivity was further demonstrated for each surrogate peptide by virtue of a BLAST search which identified peptides with the largest sequence coverage for all known isoforms of the allergens. To validate, multiple soybean varieties were analyzed by LC–MS/MS. The peak areas of the quantitative peptide monitored in the method, and MS/MS ions for one more peptide identified during method development were compared producing a ratio that should be similar across the different soybean varieties to support the quantitative peptide being monitored is specific to the endogenous soybean allergen of interest. As no reference standards were available, matrix effects on the confirmation peptide could not be monitored; thus, the method demonstrates specificity to the quantitation peptide if the ratio of peak areas for the quantitation and confirmation peptides were ±20% across soybean varieties. The ratio difference of quantitation peptide to confirmation peptide was ≤10% across soybean varieties for evaluated endogenous soybean allergens, further supporting method selectivity (Table 6).

The accuracy, precision, and ruggedness were assessed across multiple days and found to be adequate. In a multiplex assay, there may be the possibility of varying limits of detection based on the sensitivity of analytes in addition to complexities such as

Table 6. Surrogate Peptide Specificity to Respective Allergens Confirmation

protein	quantitation peptide ^a	confirmation peptide ^b	response ratio	
			soybean variety 1	soybean variety 2
Gly m 1	SYPSNATCPR	NLQLILNSCGR	1.7	1.5
Gly m 3	YMVIQGEPAVIR	N/A ^c	N/A ^c	N/A ^c
Gly m 4	ALVTDADNVIPK	SVENVEGNGGPGTIK	12.1	12.2
Gly m 5	NILEASYDTK	FEEINK	1.3	1.3
glycinin G1	VLIVPQNPFVVAAR	EQPQQNECQIQK	42.1	42.5
glycinin G2	VTAPAMR	EAFGVNMQIVR	1.4	1.4
glycinin G3	NNNPFSFLVPPK	VFDGELQEGQVLIVPQNFAVAA	0.4	0.5
glycinin G4	VESEGGLIQTWNSQHPCLK	GALGVAIPGCPETFEEPQECSNR	3.1	3.4
glycinin precursor	NGLHLPSPYPR	N/A ^c	N/A ^c	N/A ^c
Kunitz trypsin inhibitor 1	GGGIEVDSTGK	EGLQAVK	2.5	2.5
Kunitz trypsin inhibitor 3	GIGTISSPYR	LVVSK	0.5	0.5
Gly m Bd 28 K	NKPQFLAGAASLLR	TVVEEIFSK	0.4	0.4
Gly m Bd 30 K	GVITQVK	EQYSCDHPASWDWR	5.9	5.6
Gly m 8	IMENQSEELK	ELINLATMCR	0.2	0.1

^aQuantitation peptide = peptide sequence monitored in the analytical method. ^bConfirmation peptide = peptide sequence identified during method development used to support peptide specificity. ^cSuitable confirmation peptide not identified.

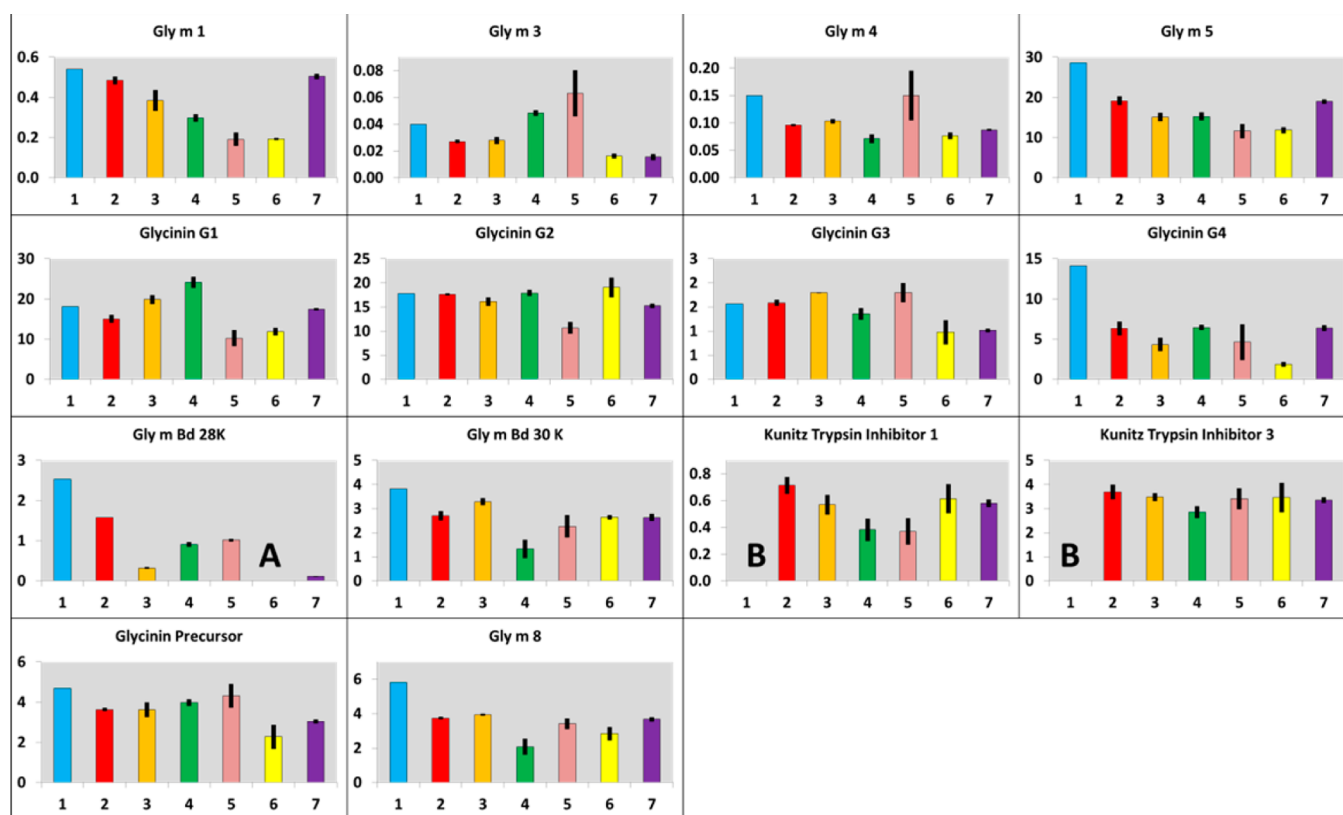


Figure 4. Plots represent interlaboratory and intralaboratory precision assessment of the multiplex assay with three biological replicate soybean samples from one field study sample measured across six participating laboratories (lines 2–7, *x*-axis) reported as micrograms of protein per milligram of dry tissue weight (*y*-axis). Laboratory 1 represents data collected by Dow AgroSciences for the field study. Symbol A represents sensitivity below limit of detection for laboratory 6. Symbol B represents surrogate peptide not monitored due alternative assay employed during compositional analysis. The detected allergen levels are comparable across laboratories with higher intralaboratory precision ($RSD < 10\%$ for all laboratories but one, which had $RSD < 20\%$).

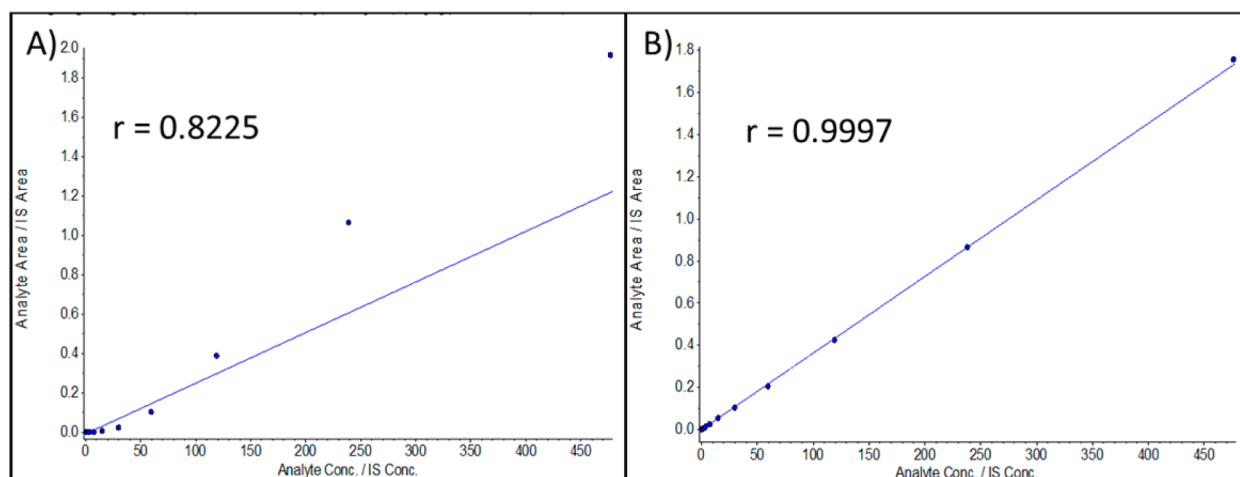


Figure 5. Higher variability and lower sensitivity were observed for hydrophobic surrogate peptides representing Gly m Bd 28 K and glycinin G4 soybean allergens. (A) Hydrophobic Gly m Bd 28 K surrogate peptide “NKPQFLAGAASLLR” loss of linearity due to adherence to labware. (B) Gly m Bd 28 K surrogate peptide “NKPQFLAGAASLLR” prepared in buffer containing 0.1% bovine serum albumin with low binding microcentrifuge tubes, resulting in improved linearity and increased sensitivity.

extraction and digestion, which would be specific to each analyte. Due to this, fortifications representing the bottom four standards were used to embody the LOQ for each endogenous soybean allergen (Figure 2). Fortifications below the limit of detection for a particular analyte were ignored. The target LOQ was empirically determined during method development (linear

response and characterized allergen expression levels) and verified by statistical approaches using the standard deviation from recovery results of the respective proposed LOQ for each endogenous soybean allergen.³⁰ The recovery equation referenced in the Methods section normalized results to peptide fortification level rather than simply baseline subtraction to

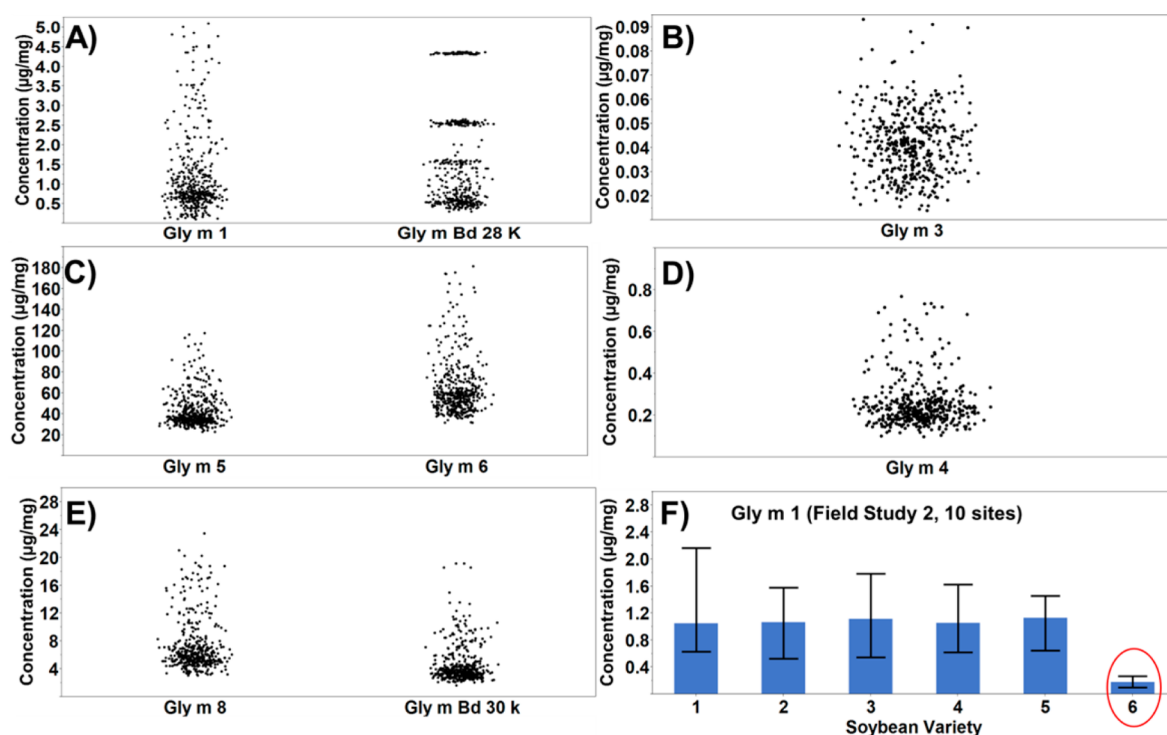


Figure 6. (A–E) Plots represent endogenous soybean allergen levels measured from commercially available non-GE soybean varieties harvested from three soybean field studies in 2011–2015. The multiplex LC–MS/MS method measures 14 individual proteins, including Gly m 1, Gly m 3, Gly m 4, Gly m 5, Gly m 6 (five individual subunits), two Kunitz trypsin inhibitors, Gly m Bd 28 K, Gly m Bd 30 K, and Gly m 8. Trypsin inhibitor was measured by enzymatic assay for composition analysis, and the five subunits of Gly m 6 were added together for total Gly m 6 abundance for each sample. Each study collected and analyzed 4 replicates per soybean variety with 3 non-GE soybean varieties present at each site across 9 or 10 sites depending on the study. Panel F demonstrates the accuracy of the assay with soybean variety 6 consistently lower than varieties 1–5.

account for the large variation between abundance in baseline measurement between peptides in the multiplexing assay (particularly at the LOQ). For example, a peptide fortification of 1.95 ng/mL would be more distinguishable for the lower expressing Gly m 1 with a mean baseline of 11 (± 0.55) ng/mL, assuming a theoretical 5% standard deviation of triplicate baseline injections, than Gly m 5 with a baseline of 275 (± 13.75) ng/mL. Both interday and intraday analysis were acceptable for 3 separate analysis with accuracy of recovery between 70 and 120% and RSD < 20% for all endogenous allergens.

Linear regression analysis with $1/y$ weighting was used to describe the detector response as a function of the calibration standard concentrations for all endogenous soybean allergens. The correlation coefficients (r) were greater than or equal to 0.995 for all of the linear regression analysis calibration curve determinations. The results indicate linearity of the detector response as a function of the standard concentration.

Refrigerated storage stability (temperature of autosampler) was evaluated over the course of three days to incorporate instrument failure over a normal weekend analysis. The results indicate that the digested peptides representing each endogenous soybean allergen were stable for at least three days when stored under refrigerated conditions (same criteria used for initial recovery experiments). Frozen storage stability of extracted supernatant as well as digested extract may also be investigated to add flexibility to high throughput analysis.

Ring Trial. A ring trial was conducted at six participating laboratories to assess precision and transferability of the multiplex endogenous soybean allergen assay. Spike recovery was performed at each laboratory following the protocol

described in the [Validation](#) section. All laboratories demonstrated that the assay had been correctly transferred with accuracy 70–120% and RSD < 20% from peptide fortification experiments with the exception of Gly m Bd 28 K and glycinin G4, which had larger variability at 3 of the laboratories due to loss of linearity of reference standards.

The results for the interlaboratory assessment of the method can be seen in [Figure 4](#). The results represent individual samples analyzed across the six participating laboratories with different LC–MS instrumentation (Agilent 6460, AB Sciex 6500, AB Sciex 4000). In general, the detected allergen levels were comparable across laboratories with higher intralaboratory precision (RSD < 10% for all laboratories but one, which had RSD < 20%). The common source of variation in absolute allergen levels between laboratories was attributed to extraction application and differences in labware. This is exemplified with the consistency of glycinin G2 across laboratories compared to Gly m Bd 28 K. The Gly m Bd 28 K allergen exhibited the greatest variance due to the hydrophobic surrogate peptide adhering to labware in the calibration curve ([Figure 5](#)). Additionally, the differences in water bath sonication resulted in variability between laboratories. Following the ring trial, the method was optimized to use a thermomixer instead of a water bath to achieve a more consistent extraction and standard curve preparation with 0.1% bovine serum albumin to prevent hydrophobic peptide loss in the calibration curve.

Field Trials. Three soybean field studies were harvested between the 2011–2015 growing seasons for compositional analysis, which included 21 commercially available non-GE soybean varieties grown across 9 or 10 field sites (depending on study) located within the major soybean-producing regions of

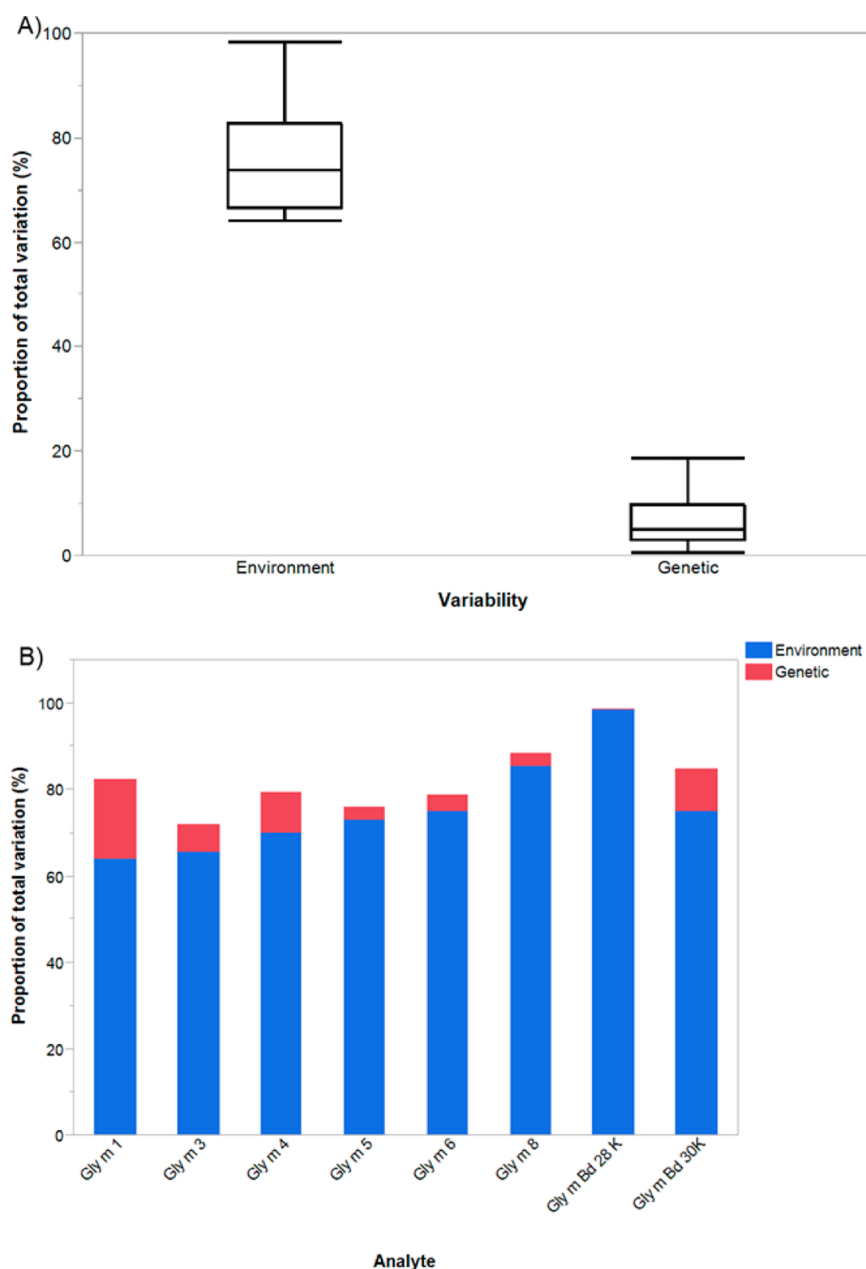


Figure 7. Endogenous soybean allergens Gly m 1, Gly m 3, Gly m 4, Gly m 5, Gly m 6 (5 individual subunits), Gly m Bd 28 K, Gly m Bd 30 K, and Gly m 8 in 21 non-GE soybean varieties measured by LC–MS/MS from 3 field studies harvested over several years from 9 to 10 field sites. Contributions of genetic and environmental factors to variation of endogenous soybean allergen abundance were evaluated using variance component analysis. The results are shown as (A) distribution of variance estimates for environmental and genetic factors for all tested allergens and (B) proportion of variation from genetic and environmental factors for each individual allergen.

the United States. The field sites represented diverse agronomic practices and environmental conditions. Endogenous soybean allergen levels were quantitatively measured by tandem mass spectrometry following the validated multiplex analytical method and compared across non-GE soybean varieties. The two Kunitz trypsin inhibitors in the multiplex method were not monitored by LC–MS/MS as they were measured by enzymatic assay in the compositional analysis.^{31,32} Gly m 6 is generally hexameric; however, the individual subunits can be found as both intact precursor proteins or covalently associated.³³ Hence, peptides specific to each subunit of Gly m 6 (glycinin G1, glycinin G2, glycinin G3, glycinin G4, and glycinin precursor) were developed in the final multiplex assay. During sample analysis, surrogate peptides for each subunit were measured, and the converted

protein results for the individual subunits of Gly m 6 were combined to obtain a total value for Gly m 6.

The results for Gly m 1 (0.10–5.10 $\mu\text{g}/\text{mg}$), Gly m 3 (0.01–0.09 $\mu\text{g}/\text{mg}$), Gly m 4 (0.10–0.77 $\mu\text{g}/\text{mg}$), Gly m 5 (22.94–117.47 $\mu\text{g}/\text{mg}$), Gly m 6 (31.79–181.33 $\mu\text{g}/\text{mg}$), Gly m Bd 28 K (0.30–4.37 $\mu\text{g}/\text{mg}$), Gly m Bd 30 K (1.61–19.15 $\mu\text{g}/\text{mg}$), and Gly m 8 (3.06–23.47 $\mu\text{g}/\text{mg}$) allergen levels in non-GE soybean varieties for individual samples over multiple growing seasons and environmental conditions are shown in Figures 6 A–E. Variance component analysis was conducted for each allergen to evaluate the contribution of genetic or environmental factors to the observed variation in soybean allergen levels. The results from variance component analysis showed that environmental factors constitute the majority of total variation (76% on average)

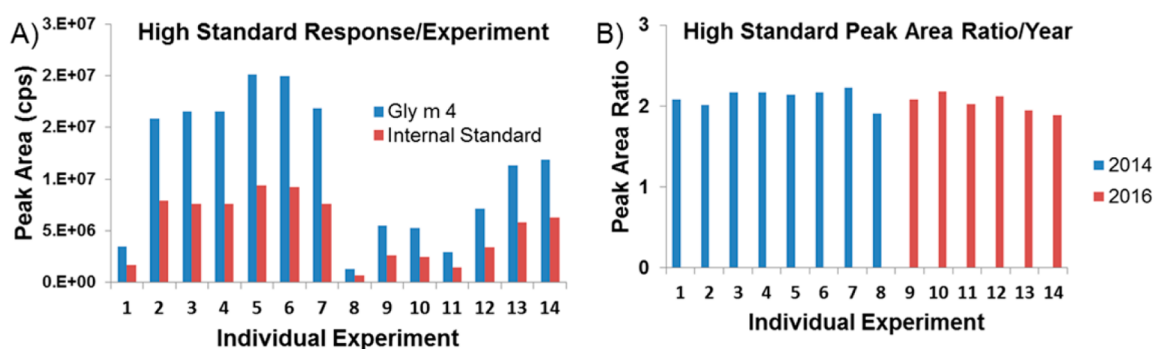


Figure 8. (A) Gly m 4 high calibration standard response for synthetic natural abundance and heavy isotope labeled peptides over 14 analysis spanning 2014–2016. The range of sensitivity was due to different instrumentation and normal fluctuation between analyses. (B) The data are normalized for accurate interpretation across experiments due to the use of heavy isotope labeled peptide internal standards and the corresponding peak area ratio.

of soybean allergen levels, while genetic factors contribute only 7% of total variation on average (Figure 7a). The breakdown of variance components by each individual allergen confirmed that environmental factors contribute more to the total variability of allergen abundance than genetic factors (Figure 7b).

A total of 14 individual allergen analyses were performed on the previously mentioned field studies incorporating different soybean cultivars, location, and growing seasons for the observed abundance of each allergen. Gly m 4 was singled out to demonstrate the value of heavy isotope internal standards to normalize samples and calibration standards. The normal range of sensitivity can be observed between instrumentation and analysis dates (Figure 8a); however, the normalized value for the high standard (Figure 8b) is consistent across analyses, allowing for correct interpretation of results. This method aspect is incredibly useful for interpretation over time across different studies or laboratories. Additionally, one of the non-GE soybean varieties consistently expressed Gly m 1 lower than other non-GE varieties across field sites in one of the studies (Figure 6F). This difference may not have clinical relevance to allergy; however, this observation demonstrates the assay's ability to accurately detect subtle variations in endogenous protein expression.

DISCUSSION

The measurement of endogenous soybean allergens is required by the European Commission (European Commission Implementing Regulation 503/2013) as part of the compositional analysis of GE crops that are common sources of food allergens. The requirement instructs that the statistical significance of any observed differences should be assessed in the context of the range of natural variations to determine biological significance. To support this regulatory requirement, the natural variation of allergen levels in comparators and conventional lines needs to be well-understood for a useful interpretation.

This research describes the development and application of a robust multiplex analytical method for measuring endogenous soybean allergen levels in soybean seed. The peptides identified through consensus sequence generation to cover all known isoforms were thoroughly validated to accurately measure Gly m 1, Gly m 3, Gly m 4, Gly m 5, Gly m 6 (five individual subunits), two Kunitz trypsin inhibitors, Gly m Bd 28 K, Gly m Bd 30 K, and Gly m 8 from soybean seed in a high throughput format. The versatility of the method to monitor multiple proteins from a single extraction (14 individual proteins when considering the 5 individual subunits of Gly m 6 incorporating a large range of protein chemistries) as well as having no need to purify protein

standards demonstrates the advantage of the technology for allergen quantitation. Furthermore, this is the first study to demonstrate assay precision and robustness across multiple independent laboratories for endogenous allergen quantitation. The agreement of results collected across the six participating laboratories shows that the assay can be rapidly adapted with minimal optimization in multiple laboratories to yield robust and reproducible results for endogenous analyte quantitation.

Although the method was validated to accurately and reproducibly measure peptides as surrogates to respective allergens for risk assessment, the accuracy to target protein levels is based on the assumption that enzymatic digestion is followed to completion, which was validated with a timecourse digestion. Future work should evaluate intact purified protein standards in the context of absolute allergen abundance. The soybean allergen levels measured by this LC–MS/MS method (Gly m 4, 0.10–0.77 $\mu\text{g}/\text{mg}$; Gly m 5, 22.94–117.47 $\mu\text{g}/\text{mg}$; Gly m 6, 31.79–181.33 $\mu\text{g}/\text{mg}$; Gly m Bd 28 K, 0.3–4.37 $\mu\text{g}/\text{mg}$; Gly m Bd 30 K, 1.61–19.15 $\mu\text{g}/\text{mg}$) are comparable to levels reported using ELISA¹⁸ (Gly m 4, 0.02–0.33 $\mu\text{g}/\text{mg}$; Gly m 5, 5.52–90.40 $\mu\text{g}/\text{mg}$; Gly m 6, 78.40–393 $\mu\text{g}/\text{mg}$; Gly m Bd 28 K, 0.1–0.47 $\mu\text{g}/\text{mg}$; Gly m Bd 30 K, 0.3–3.49 $\mu\text{g}/\text{mg}$) and 2DLC–UV/MS¹⁷ (Gly m 4, 0.36–0.61 $\mu\text{g}/\text{mg}$), providing further confidence in the accuracy of the methodology.

The multiplexing method was further applied to three field studies spanning multiple years and environmental conditions to provide an understanding of the range of the endogenous soybean allergen levels from non-GE soybean varieties. The levels of endogenous soybean allergens varied from 5-fold in Gly m 6 to 50-fold in Gly m 1 from 624 individual conventional soybean seed samples analyzed. Additionally, the results are consistent with previously reported studies demonstrating variations in soybean allergen levels are affected more from environmental factors than genetic factors.^{18,34} The dominant effect of environment on soybean allergen levels, the observed natural range of abundances of respective allergens, and large contribution of Gly m 5 and Gly m 6 allergen levels to the total allergen profile calls into question the value of uniquely monitoring the genotypic effects of GE soybeans, especially when individuals with soybean allergy are typically instructed to avoid all food containing soybean-derived ingredients. While the utility of endogenous allergen assessment for GE crops is still highly debated,^{35–41} the LC–MS/MS approach is useful for the accurate measurement of endogenous allergens in soybean seed.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.7b01018.

Consensus protein sequence selection for all endogenous soybean allergens and SDS-PAGE gel performed for digestion efficiency (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rhill1@dow.com.

ORCID

Ryan C. Hill: 0000-0002-8424-5341

Notes

The authors declare the following competing financial interest(s): R.C.H., T.J.O., X.W., R.A.H., G.S., and B.S. are employees of Dow AgroSciences LLC, a wholly owned subsidiary of The Dow Chemical Company, which develops transgenic crops and produces insecticides, herbicides, and fungicides for agricultural applications. K.C., J.S., B.M., L.S., A.X., and J.B. are employed by EAG Laboratories, Covance Laboratories, Critical Path Services, EPL Bioanalytical Services, Primera Analytical Solutions, and Pyxant Laboratories, respectively, which performed the independent method assessment under contract from Dow AgroSciences LLC.

■ ACKNOWLEDGMENTS

We thank Ping Song and Jeff Gilbert for helpful comments and discussions. The multiplex technology is patented by The Dow Chemical Company (US 8,227,252, WO 2016/025516) and commercialized as Plextein Trademark of The Dow Chemical Company ("Dow") or an affiliated company of Dow. The 2DLC-UV/MS technology is patented by The Dow Chemical Company (US9170231).

■ REFERENCES

- (1) Natarajan, S.; Luthria, D.; Hanhong, B.; Lakshman, D.; Mitra, A. Transgenic soybeans and soybean protein analysis: An overview. *J. Agric. Food Chem.* **2013**, *61*, 11736–11743.
- (2) Taylor, S. L.; Hefle, S. L. Will Genetically Modified Foods be Allergenic? *J. Allergy Clin. Immunol.* **2001**, *107* (5), 765–771.
- (3) Arpaia, S. Scientific Opinion on application (EFSA-GMO-UK-2009–76) for the placing on the market of soybean MON 87769 genetically modified to contain stearidonic acid, for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Monsanto. *EFSA J.* **2014**, *12*, 1–41.
- (4) Arpaia, S. Scientific Opinion on application EFSA-GMO-NL-2007–45 for the placing on the market of herbicide-tolerant, high-oleic acid, genetically modified soybean 305423 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Pioneer. *EFSA J.* **2013**, *11*, 1–35.
- (5) Birch, A. N. Scientific Opinion on an application by Dow AgroSciences (EFSA-GMO-NL-2013–116) for placing on the market of genetically modified insect-resistant soybean DAS-81419–2 for food and feed uses, import and processing under Regulation (EC) No 1829/2003. *EFSA J.* **2016**, *14*, 1–35.
- (6) Nicolai, A.; Manzo, A.; Veronesi, F.; Rosellini, D. An Overview of the last 10 years of genetically engineered crop safety research. *Crit. Rev. Biotechnol.* **2014**, *34*, 77–88.
- (7) CODEX. *Foods Derived from Biotechnology*; Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme: Rome, 2009.
- (8) EFSA. Commission Implementing Regulation (EU) No. 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No. 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No. 1981/2006. *Off. J. Eur. Union*, **2013**, L157, 148.
- (9) Andersson, H. C. Scientific opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. *EFSA J.* **2010**, *8*, 1–168.
- (10) Andersson, H. C. Panel on Genetically Modified Organisms (GMO): Guidance for risk assessment of food and feed from genetically modified plants. *EFSA J.* **2011**, *9* (5), 2150.
- (11) EFSA. Workshop on key allergens and compositional analysis in the allergenicity assessment of genetically modified plants. *EFSA J.*, **2012**, 289.
- (12) OECD. *Revised Consensus Document on Compositional Considerations for New Varieties of Soybean*; Organisation for Economic Co-operation and Development: 2012.
- (13) EFSA. Draft Guidance on Allergenicity Assessment of Genetically Modified Plants. Public Consultation. *EFSA J.*, **2016**.
- (14) Ladics, G. S.; Budziszewski, G. J.; Herman, R. A.; Herouet-Guicheney, C.; Joshi, S.; Lipscomb, E. A.; McClain, S.; Ward, J. M. Measurement of endogenous allergens in genetically modified soybeans – short communication. *Regul. Toxicol. Pharmacol.* **2014**, *70* (1), 75–79.
- (15) Panda, R.; Ariyaratna, H.; Amnuaycheewa, P.; Tetteh, A.; Pramod, S. N.; Taylor, S. L.; Ballmer-Weber, B. K.; Goodman, R. E. Challenges in testing genetically modified crops for potential increases in endogenous allergen expression for safety. *Allergy* **2013**, *68*, 142–151.
- (16) Geng, T.; Liu, K.; Frazier, R.; Shi, L.; Bell, E.; Glenn, K.; Ward, J. M. Development of a Sandwich ELISA for Quantification of Gly m 4, a Soybean Allergen. *J. Agric. Food Chem.* **2015**, *63*, 4947–4953.
- (17) Julka, S.; Kuppannan, K.; Karnoup, A.; Dielman, D.; Schafer, B.; Young, S. A. Quantification of Gly m 4 Protein, A major Soybean Allergen, By Two-Dimensional Liquid Chromatography with Ultra-violet and Mass Spectrometry Detection. *Anal. Chem.* **2012**, *84*, 10019–10030.
- (18) Geng, T.; Stojin, T.; Liu, K.; Schaale, B.; Postin, C.; Ward, J. M.; Wang, Y.; Liu, Z. L.; Li, B.; Glenn, K. C. Natural variability of allergen levels in conventional soybeans: assessing variation across North and South America from five production years. *J. Agric. Food Chem.* **2017**, *65*, 463.
- (19) Houston, N. L.; Lee, D.; Stevenson, S. E.; Ladics, G. S.; Bannon, G. A.; McClain, S.; Privalle, L.; Stagg, N.; Herouet-Guicheney, C.; MacIntosh, S. C.; Thelen, J. J. Quantitation of Soybean Allergens Using Tandem Mass Spectrometry. *J. Prot. Research* **2011**, *10*, 763–773.
- (20) Settlege, S. B.; Eble, J. E.; Bhanushali, J. K.; Cheever, M. L.; Gao, A.; Goldstrohm, D. A.; Hill, R.; Hu, T. X.; Powley, C. R.; Unger, A.; Shan, G. Validation Parameters for Quantitating Specific Proteins Using ELISA or LC-MS/MS: Survey Results. *Food Anal. Methods* **2017**, *10*, 1339.
- (21) U.S. Department of Health and Human Services Food and Drug Administration. *Guidance for Industry: Bioanalytical Method Validation*; Washington, D.C., 2001.
- (22) Hill, R. C.; Oman, T. O.; Shan, G.; Schafer, B.; Eble, J.; Chen, C. Development and Validation of a Multiplexed Protein Quantitation Assay for the Determination of Three Recombinant Proteins in Soybean Tissues by Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS). *J. Agric. Food Chem.* **2015**, *63*, 7450–7461.
- (23) Skogerson, K.; Harrigan, G. G.; Reynolds, T. L.; Halls, S. C.; Ruebelt, M.; Iandolino, A.; Pandravada, A.; Glenn, K. C.; Fiehn, O. Impact of Genetics and Environment on the Metabolite Composition of Maize Grain. *J. Agric. Food Chem.* **2010**, *58*, 3600–3610.
- (24) Zhou, J.; Berman, K. H.; Breeze, M. L.; Nemeth, M. A.; Oliveira, W. S.; Braga, D. P.; Berger, G. U.; Harrigan, G. G. Compositional variability in conventional and glyphosate tolerant soybean (*Glycine max* L.) varieties grown in different regions in Brazil. *J. Agric. Food Chem.* **2011**, *59*, 11652–11656.
- (25) Zhou, J.; Harrigan, G. G.; Berman, K. H.; Webb, E. G.; Klusmeyer, T. H.; Nemeth, M. A. Stability in the composition equivalence of grain from insect-protected maize and seed from glyphosate-tolerant soybean

to conventional counterparts over multiple seasons, locations, and breeding germplasms. *J. Agric. Food Chem.* **2011**, *59*, 8822–8828.

(26) Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. M. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal. Biochem.* **2005**, *342*, 214–220.

(27) Bronsema, K. J.; Bischoff, R.; van de Merbel, N. C. High-Sensitivity LC–MS/MS Quantification of Peptides and Proteins in Complex Biological Samples: The Impact of Enzymatic Digestion and Internal Standard Selection on Method Performance. *Anal. Chem.* **2013**, *85*, 9528–9535.

(28) Lowenthal, M. S.; Liang, Y.; Phinney, K.; Stein, S. E. Quantitative Bottom-Up Proteomics Depends on Digestion Conditions. *Anal. Chem.* **2014**, *86* (1), 551–558.

(29) Jenkins, R.; Duggan, J. X.; Aubry, A.; Zeng, J.; Lee, J. W.; Cojocaru, L.; Dufield, D.; Garofolo, F.; Kaur, S.; Schultz, G. A.; Xu, K.; Yang, Z.; Yu, J.; Zhang, Y. J.; Vazvaei, F. Recommendations for Validation of LC–MS/MS Bioanalytical Methods for Protein Biotherapeutics. *AAPS J.* **2015**, *17*, 1–16.

(30) Keith, L. H.; Crummett, W.; Deegan, J.; Libby, R. A.; Taylor, J. K.; Wentler, G. Principles of environmental analysis. *Anal. Chem.* **1983**, *55*, 2210–2218.

(31) Lepping, M. D.; Herman, R. A.; Potts, B. L. Compositional Equivalence of DAS-44406–6 (AAD-12 + 2mEPS + PAT) Herbicide-Tolerant Soybean and Nontransgenic Soybean. *J. Agric. Food Chem.* **2013**, *61*, 11180–11190.

(32) Fast, B. J.; Schafer, A. C.; Johnson, T. Y.; Potts, B. L. Insect-Protected Event DAS-81419–2 Soybean (*Glycine max* L.) Grown in the United States and Brazil Is Compositionally Equivalent to Non-transgenic Soybean. *J. Agric. Food Chem.* **2015**, *63* (7), 2063–2073.

(33) Wang, F.; Robotham, J. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Ana o 2, a Major Cashew (*Anacardium occidentale* L.) Nut Allergen of the Legumin Family. *Int. Arch. Allergy Immunol.* **2003**, *132* (1), 27–39.

(34) Stevenson, S. E.; Woods, C. A.; Hong, B.; Kong, X.; Thelen, J. J.; Ladics, G. S. Environmental effects on allergen levels in commercially grown non-genetically modified soybeans: assessing variation across North America. *Front. Plant Sci.* **2012**, *196* (3), 1–13.

(35) Herman, R. A.; Ladics, G. S. Endogenous allergen upregulation: Transgenic vs. traditionally bred crops. *Food Chem. Toxicol.* **2011**, *49*, 2667–2669.

(36) Graf, L.; Hayder, H.; Mueller, U. Endogenous allergens in the regulatory assessment of genetically engineered crops. *Food Chem. Toxicol.* **2014**, *73*, 17–20.

(37) Goodman, R. E.; Vieths, S.; Sampson, H. A.; Hill, D.; Ebisawa, M.; Taylor, S. L.; van Ree, R. Allergenicity assessment of genetically modified crops—what makes sense? *Nat. Biotechnol.* **2008**, *26*, 73–81.

(38) Goodman, R. E.; Panda, R.; Ariyaratna, H. Evaluation of endogenous allergens for the safety evaluation of genetically engineered food crops: review of potential risks, test methods, examples and relevance. *J. Agric. Food Chem.* **2013**, *61*, 8317–8332.

(39) Doerr, N.; Ladics, G.; McClain, S.; Herouet-Guicheney, C.; Poulsen, L. K.; Privalle, L.; Stagg, N. Evaluating biological variation in non-transgenic crops: Executive summary from the ILSI Health and Environmental Sciences Institute Workshop, November 16–17, 2009, Paris, France. *Regul. Toxicol. Pharmacol.* **2010**, *58*, S2–S7.

(40) Fernandez, A.; Mills, E. N. C.; Lovik, M.; Spoek, A.; Germini, A.; Mikalsen, A.; Wal, J. M. Endogenous allergens and compositional analysis in the allergenicity assessment of genetically modified plants. *Food Chem. Toxicol.* **2013**, *62*, 1–6.

(41) Selb, R.; Wal, J. M.; Moreno, F. J.; Lovik, M.; Mills, C.; Hoffmann-Sommergruber, K.; Fernandez, A. Assessment of endogenous allergenicity of genetically modified plants exemplified by soybean – Where do we stand? *Food Chem. Toxicol.* **2017**, *101*, 139–148.

Title: Development, Validation, and Inter-laboratory Evaluation of a Quantitative Multiplexing Method to assess levels of Ten Endogenous Allergens in Soybean Seed and its Application to Field Trials Spanning Three Growing Seasons

Authors: Ryan C. Hill^{1*}, Trent J. Oman¹, Xiujuan Wang¹, Guomin Shan¹, Barry Schafer¹, Rod A. Herman¹, Rowel Tobias², Jeff Shippar³, Bhaskar Malayappan⁴, Li Sheng⁵, Austin Xu⁶, Jason Bradshaw⁷

¹Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268, United States

²EAG Laboratories, 4780 Discovery Drive, Columbia, MO, 65201, United States

³Covance Laboratories, 3301 Kinsman Blvd., Madison, WI, 53704, United States

⁴Critical Path Services LLC, 3070 McCann Farm Drive, Garnet Valley, PA, 19060, United States

⁵EPL Bioanalytical Services, 9095 W. Harristown Blvd, Niantic, IL 62551, United States

⁶Primera Analytical Solutions, 259 Wall St, Princeton, NJ 08540, United States

⁷Pyxant Labs, 4720 Forge Rd #106, Colorado Springs, CO, 80907, United States

Supporting Data Contents:

Supporting Information S1-S14: Consensus Sequence Generation

Supporting Information S15: Confirmatory SDS-PAGE gel for digestion completeness.

Supporting Information S1: Gly m 1 Consensus Sequence Generation

Gly m 1 Protein Sequences:

>Gly m 1 AAB34755.1 42aa

ALITRPSXPDLISIXLNILGGSGLGTVDXXALIGGLXDXXAIV

>Gly m 1 ABA54898.1 119aa

MGSKVVASVALLLSINILFISMVSSSSHYDPQPQPSHVTALITRPSCPDLSICLNILGGSGLGTVDCCAL
IGGLGDIEAIVCLCIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT

>Gly m 1 ABA54897.1 134aa

MGSKVVASVALLLSINILFISMVSSSSHYDPPPPCYVPAPLTPPPSLSPPPSPGSPCDLSVC
LNILDGSPADDCCALIADLVLEASVCLCIQLRVLGIVNLDLNLQLILNACGPSYPSNATCPRT

>Gly m 1 ABA54899 118aa

GSKVVASVALLLSINILFISMVSSSSHYDPQPQPSHVTALITRPSCPDLSICLNILGGSGLGTVDCCALI
GGLGDIEAIVCLCIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT

>Gly m 1 Glyma15g13740.1 120aa

MGSKVVAYVALLLSINILFISMVSSSSHYDPQPQPSYVTALITRPSCPDLSVCLNILGGYLGTVDDCCAL
IGGLGDIEATVCLCIQLRALGILNLRNLQLILNACGPSYPSNATCPRT*

>Gly m 1 Glyma15g13770.1 130aa

MGSKVVASVALLLSINILFISMVSSSSHYDPPPPCYVPAPFTPPPSLSPPPSPGSPCDLSVCLNILD
GSPADDCCALIADLVLEASVCLCIQLRVLGIVNLDLNLQLILNACGPSYPSNATCPRT*

>Gly m 1 Glyma15g13750.1 120aa

MGSKVVASVALLLSINILFISMVSSSSHYDPPPPQPSYVTALITRPSCPDLSICLNILGGSGLGTVDCCAL
IGGLGDIEAIVCLCIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT*

Gly m 1 Protein Sequence Alignments:

	1		50
Gly m 1 AAB34755.1	(1)	-----ALITRPSX---	
Gly m 1 ABA54898.1	(1)	MGSKVVASVALLLSINILFISMVSSSSHYDPQPQPSHVTALITRPS---	
Gly m 1 ABA54897.1	(1)	MGSKVVASVALLLSINILFISMVSSSSHYDPPPPCYVPAPLTPPPSLSP	
Gly m 1 ABA54899	(1)	GSKVVASVALLLSINILFISMVSSSSHYDPQPQPSHVTALITRPS----	
Gly m 1 Glyma15g13740.1	(1)	MGSKVVA ~~Y~~ VALLLSINILFISMVSSSSHYDPQPQPSYVTALITRPS----	
Gly m 1 Glyma15g13770.1	(1)	MGSKVVASVALLLSINILFISMVSSSSHYDPPPPCYVPAPFTPPP----	
Gly m 1 Glyma15g13750.1	(1)	MGSKVVASVALLLSINILFISMVSSSSHYDPPQPSYVTALITRPS----	
Consensus	(1)	MGSKVVASVALLLSINILFISMVSSSSHYDPQPQPSYVTALITRPS	
	51		100
Gly m 1 AAB34755.1	(9)	-----PDLISIXLNILGGSGLGTVDXXALIGGLXDXXAIV--	
Gly m 1 ABA54898.1	(48)	-----PDLISICLNILGGSGLGTVDCCALIGGLGDIEAIVCL	
Gly m 1 ABA54897.1	(51)	PPSLSPPPSPGSPCDLSVCLNILDG--SPADDCCALIADLVLEASVCL	
Gly m 1 ABA54899	(46)	-----CPDLSICLNILGGSGLGTVDCCALIGGLGDIEAIVCL	
Gly m 1 Glyma15g13740.1	(47)	-----CPDLSVCLNILGGYLGTVDDCCALIGGLGDIEATVCL	
Gly m 1 Glyma15g13770.1	(47)	-PSLSPPPSPGSPCDLSVCLNILDG--SPADDCCALIADLVLEASVCL	
Gly m 1 Glyma15g13750.1	(47)	-----CPDLSICLNILGGSGLGTVDCCALIGGLGDIEAIVCL	
Consensus	(51)	CPDLSICLNILGGSGLGTVDCCALIGGLGDIEAIVCL	
	101		137
Gly m 1 AAB34755.1	(43)	-----	
Gly m 1 ABA54898.1	(84)	CIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT-	
Gly m 1 ABA54897.1	(99)	CIQLRVLGIVNLDLNLQLILNACGPSYPSNATCPRT-	
Gly m 1 ABA54899	(83)	CIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT-	
Gly m 1 Glyma15g13740.1	(84)	CIQLRALGILNLRNLQLILNACGPSYPSNATCPRT-	
Gly m 1 Glyma15g13770.1	(94)	CIQLRVLGIVNLDLNLQLILNACGPSYPSNATCPRT-	
Gly m 1 Glyma15g13750.1	(84)	CIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT-	
Consensus	(101)	CIQLRALGILNLRNLQLILNACGRSYPSNATCPRT	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

*Gly m 1 Gly m 1 Ping ABA54898.1 119aa

MGSKVVASVALLLSINILFISMVSSSSHYDPQPQPSHVTALITRPSCPDLSICLNILGGSGLGTVDCCAL
IGGLGDIEAIVCLCIQLRALGILNLRNLQLILNSCGRSYPSNATCPRT

Supporting Information S2: Gly m 3 Consensus Sequence Generation

Gly m 3 Protein Sequences:

>Gly m 3 ABU97472.1 131aa

MSWQAYVDDHLLCEIEGNHLTHAAIIGQDGSVWAQSTNFPQFKPEEITAINNDFNEPGSLAPTGLYIGGT
KYMVIQGEPGA VIRGKKGPGGVTVKKTGAALIIIGIYDEPMTPGQCNMVVERLGDYLIDQGL

>Gly m 3 CAA11755.1 131aa

MSWQAYVDDHLLCGIEGNHLTHAAIIGQDGSVWLQSTDFPQFKPEEITAIMNDFNEPGSLAPTGLYLGGT
KYMVIQGEPGA VIRGKKGPGGVTVKKTGAALIIIGIYDEPMTPGQCNMVVERLGDYLIDQGY

>Gly m 3 O65809.1 131aa

MSWQAYVDDHLLCDIEGNHLTHAAIIGQDGSVWAQSTDFPQFKPEEITAIMNDFNEPGSLAPTGLYLGGT
KYMVIQGEPGA VIRGKKGPGGVTVKKTGAALIIIGIYDEPMTPGQCNMVVERPGDYLIDQGY

Gly m 3 Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1		50
Gly m 3	ABU97472.1	(1)	MSWQAYVDDHLLCEIEGNHLTHAAIIGQDGSVWAQSTNFPQFKPEEITAI		
Gly m 3	CAA11755.1	(1)	MSWQAYVDDHLLCGIEGNHLTHAAIIGQDGSVWLQSTDFPQFKPEEITAI		
Gly m 3	O65809.1	(1)	MSWQAYVDDHLLCDIEGNHLTHAAIIGQDGSVWAQSTDFPQFKPEEITAI		
	Consensus	(1)	MSWQAYVDDHLLC IEGNHLTHAAIIGQDGSVWAQSTDFPQFKPEEITAI		
			51		100
Gly m 3	ABU97472.1	(51)	NNDFNEPGSLAPTGLYIGGTKYKYMVIQGEPGA VIRGKKGPGGVTVKKTGAA		
Gly m 3	CAA11755.1	(51)	NNDFNEPGSLAPTGLYLGGTKYKYMVIQGEPGA VIRGKKGPGGVTVKKTGAA		
Gly m 3	O65809.1	(51)	NNDFNEPGSLAPTGLYLGGTKYKYMVIQGEPGA VIRGKKGPGGVTVKKTGAA		
	Consensus	(51)	MNDFNEPGSLAPTGLYLGGTKYKYMVIQGEPGA VIRGKKGPGGVTVKKTGAA		
			101		131
Gly m 3	ABU97472.1	(101)	LIIGIYDEPMTPGQCNMVVERLGDYLIDQGL		
Gly m 3	CAA11755.1	(101)	LIIGIYDEPMTPGQCNMVVERLGDYLIDQGY		
Gly m 3	O65809.1	(101)	LIIGIYDEPMTPGQCNMVVERPGDYLIDQGY		
	Consensus	(101)	LIIGIYDEPMTPGQCNMVVERLGDYLIDQGY		

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

*Gly m 3 Consensus Sequence 131aa

MSWQAYVDDHLLCGIEGNHLTHAAIIGQDGSVWLQSTDFPQFKPEEITAIMNDFNEPGSLAPTGLYLGGT
KYMVIQGEPGA VIRGKKGPGGVTVKKTGAALIIIGIYDEPMTPGQCNMVVERLGDYLIDQGY

Supporting Information S3: Gly m 4 Consensus Sequence Generation

Gly m 4 Protein Sequences:

>Gly m 4 P26987 158aa

MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNNGPGTIKKITFLEDGETKFLVHKIESIDEANLGYSYSVVGGAALPDTAEKITFDSKLV
AGPNNGGSAGKLTVKYETKGAEPNQDELKTGKAKADALFKAIEAYLLAHPDYN

>Gly m 4 Glyma07g37240.1 165aa

MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNNGPGTIKKITFLEADVNEWIDGETKFLVHKIESIDEANLGYSYSVVGGAALPDTAEK
ITFDSKLVAGPNNGGSAGKLTVKYETKGAEPNQDELKTGKAKADALFKAIEAYLLAHPDYN

>Gly m 4 Glyma17g03365.1 159aa

MGIFTFEDEITSPVAPATLYKALVTDADNIIPKALDSFKSVENVEGNNGPGTIKKITFVEDGETKFLVHKIEAVDEANLGYSYSVVGGAALPDTAEKITFHSKLA
AGPNNGGSAGKLTVEYQTKGDAQPNQDQLKTGKAKADALFKAIEAYLLANPDYN*

>Gly m 4 Glyma07g37240.3 147aa

MGVFTFEDEINSPVAPATLYKALDSFKSVENVEGNNGPGTIKKITFLEDGETKFLVHKIESIDEANLGYSYSVVGGAALPDTAEKITFDSKLVAGPNNGGSAGKL
TVKYETKGAEPNQDELKTGKAKADALFKAIEAYLLAHPDYN*

>Gly m 4 Glyma07g37270.2 159aa

MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNNGPGTIKKITFLEDGETKFLVHKIEAIDEANLGYSYSVVGGDGLPDTVEKITFECKL
AAGANGGSAGKLTVKYQTKGDAQPNQDDLKIGKAKSDALFKAIEAYLLAHPDYN*

>Gly m 4 Glyma07g37270.1 159aa

MGVFTFEDETTSPVAPATLYKALVTDADNVIPKAVIDAFRSVENVEGNNGPGTIKKITFLEDGETKFLVHKIEAIDEANLGYSYSVVGGDGLPDTVEKITFECKL
AAGANGGSAGKLTVKYQTKGDAQPNQDDLKIGKAKSDALFKAIEAYLLAHPDYN*

>Gly m 4 Glyma17g03360.1 159aa

MGVFTFEDETTSPVAPATLYKALVTDADNVIPKAVIDAFRSVENLEGNGGPGTIKKITFVEDGESKFLVHKIESVDEANLGYSYSVVGGLPDTVEKITFECKL
AAGANGGSAGKLTVKYQTKGDAQPNPDDLKIGKVKSDALFKAIEAYLLANPHYN*

>Gly m 4 Glyma17g03350.1 159aa

MGIFTFEDETTSPVAPATLYKALVTDADNVIPKAVEAFRSVENLEGNGGPGTIKKITFVEDGESKFLVHKIESVDEANLGYSYSVVGGLPDTVEKITFECKL
AAGANGGSAGKLTVKYQTKGDAQPNPDDLKIGKVKSDALFKAIEAYLLANPHYN*

>Gly m 4 Glyma09g04510.1 159aa

MGVFTFEDETTSTVAPARLYKALVKDADNLVPKAVEAIKSVEIVEGNNGGPGTIKKLTFVEDGQTKYVLHKVEAIDEANWGYNYSVVGGLPDTVEKISFEA
KLVDPNNGGSIKITYKYQTKGDANPSEELKSGKAKGDALFKALEGYVLANPDYN*

>Gly m 4 Glyma15g15590.1 159aa

MGVFTFEDETTSTVAPARLYKALVKDADNLVPKAVEAIKSVEIVEGNNGGPGTIKKLTFVEDGQTKYVLHKVEAIDEANWGYNYSVVGGLPDTVEKISFEA
KLVEGASGGSIAKITVKYQTKGDVNPSEELKSGKAKGDALFKALEGYVLANPDYN*

Supporting Information S3 (Continued): Gly m 4 Consensus Sequence Generation

Gly m 4 Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

		1	50
Gly m 4 P26987	(1)	MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPF	
Gly m 4 Glyma07g37240.1	(1)	MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPF	
Gly m 4 Glyma17g03365.1	(1)	MGIFTFEDEITSPVAPATLYKALVTDADNIPKALDSFKSVENVEGNGGPF	
Gly m 4 Glyma07g37240.3	(1)	MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPF	
Gly m 4 Glyma07g37270.2	(1)	MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPF	
Gly m 4 Glyma07g37270.1	(1)	MGVFTFEDETTSPVAPATLYKALVTDADNVIPKAVDAFRSVENVEGNGGPF	
Gly m 4 Glyma17g03360.1	(1)	MGVFTFEDETTSPVAPATLYKALVTDADNVIPKAVDAFRSVENVEGNGGPF	
Gly m 4 Glyma17g03350.1	(1)	MGIFTFEDETTSPVAPATLYKALVTDADNVIPKAVEAFRSVENVEGNGGPF	
Gly m 4 Glyma09g04510.1	(1)	MGVFTFEDETTSTVAPARLYKALVKDADNLVPKAVEAISKVEIVEGNGGPF	
Gly m 4 Glyma15g15590.1	(1)	MGVFTFEDETTSTVAPARLYKALVKDADNLVPKAVEAISKVEIVEGNGGPF	
Consensus	(1)	MGVFTFEDE TSPVAPATLYKALVTDADNVIPKA D FKSVENVEGNGGPF	
		51	100
Gly m 4 P26987	(51)	GTIKKITFLED-----GETKFLVHKIESTIDEANLGYSYSVVGGAALPD	
Gly m 4 Glyma07g37240.1	(51)	GTIKKITFLEADVNEWIDGETKFLVHKIESTIDEANLGYSYSVVGGAALPD	
Gly m 4 Glyma17g03365.1	(51)	GTIKKITFVED-----GETKFLVHKIEAVDEANLGYSYSVVGGAALPD	
Gly m 4 Glyma07g37240.3	(39)	GTIKKITFLED--D-----GETKFLVHKIESTIDEANLGYSYSVVGGAALPD	
Gly m 4 Glyma07g37270.2	(51)	GTIKKITFLED-----GETKFLVHKIEAIDEANLGYSYSVVGGDGLPD	
Gly m 4 Glyma07g37270.1	(51)	GTIKKITFLED-----GETKFLVHKIEAIDEANLGYSYSVVGGDGLPD	
Gly m 4 Glyma17g03360.1	(51)	GTIKKITFVED-----GESKFLVHKIESVDEANLGYSYSVVGGVGLPD	
Gly m 4 Glyma17g03350.1	(51)	GTIKKITFVED-----GESKFLVHKIESVDEANLGYSYSVVGGVGLPD	
Gly m 4 Glyma09g04510.1	(51)	GTIKKITFVED-----GQTKYVLHKVEAIDEANWGYNYSVVGGVGLPD	
Gly m 4 Glyma15g15590.1	(51)	GTIKKITFVED-----GQTKYVLHKVEAIDEANWGYNYSVVGGVGLPD	
Consensus	(51)	GTIKKITF ED GETKFLVHKIE IDEANLGYSYSVVGGLPD	
		101	150
Gly m 4 P26987	(94)	TAEKITFDSKLVAGPNNGGSAGKLTVKYETKGDAEFNQDELKTGKAKADAL	
Gly m 4 Glyma07g37240.1	(101)	TAEKITFDSKLVAGPNNGGSAGKLTVKYETKGDAEFNQDELKTGKAKADAL	
Gly m 4 Glyma17g03365.1	(94)	TAEKITFHSKLAAAGPNNGGSAGKLTVEYQTKGDAQFNQDQLKTGKAKADAL	
Gly m 4 Glyma07g37240.3	(82)	TAEKITFDSKLVAGPNNGGSAGKLTVKYETKGDAEFNQDELKTGKAKADAL	
Gly m 4 Glyma07g37270.2	(94)	TVEKITFECKLAAGANGGSAGKLTVKYQTKGDAQFNQDDLKIGKAKSDAL	
Gly m 4 Glyma07g37270.1	(94)	TVEKITFECKLAAGANGGSAGKLTVKYQTKGDAQFNQDDLKIGKAKSDAL	
Gly m 4 Glyma17g03360.1	(94)	TVEKITFECKLAAGANGGSAGKLTVKYQTKGDAQFNPDLLKIGKVKSDAL	
Gly m 4 Glyma17g03350.1	(94)	TVEKITFECKLAAGANGGSAGKLTVKYQTKGDAQFNPDLLKIGKVKSDAL	
Gly m 4 Glyma09g04510.1	(94)	TVEKISFEAKLVADPNNGGSIAKITVKYQTKGDANFSEELKSGKAKGDAL	
Gly m 4 Glyma15g15590.1	(94)	TVEKISFEAKLVEASGGGIAKITVKYQTKGDVNFSEELKSGKAKGDAL	
Consensus	(101)	TVEKITFE KL AG NGGSAGKLTVKYQTKGDAQFNQDELK GKAK DAL	
		151	165
Gly m 4 P26987	(144)	FKAIEAYLLAHPDYN	
Gly m 4 Glyma07g37240.1	(151)	FKAIEAYLLAHPDYN	
Gly m 4 Glyma17g03365.1	(144)	FKAIEAYLLANPDYN	
Gly m 4 Glyma07g37240.3	(132)	FKAIEAYLLAHPDYN	
Gly m 4 Glyma07g37270.2	(144)	FKAIVEAYLLAHPDYN	
Gly m 4 Glyma07g37270.1	(144)	FKAIVEAYLLAHPDYN	
Gly m 4 Glyma17g03360.1	(144)	FKAIVEAYLLANPHYN	
Gly m 4 Glyma17g03350.1	(144)	FKAIVEAYLLANPHYN	
Gly m 4 Glyma09g04510.1	(144)	FKALEGYVLANPDYN	
Gly m 4 Glyma15g15590.1	(144)	FKALEGYVLANPDYN	
Consensus	(151)	FKA EAYLLA PDYN	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 4 P26987 158aa

MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPGTIKKITFLEDGETKFLVHKIESIDEANLGYSYSVVGGAALPDTAEKITFDSKLV
AGPNNGGSAGKLTVKYETKGDAPFNQDELKTGKAKADALFKAIEAYLLAHPDYN

Supporting Information S4: Gly m 5 Consensus Sequence Generation

>ABH09130

MFGIVYWEKQNP SHNKL RSCNSEKDSYRNQACHARC NLLKVEEEEECEEGQIPRPRQHPERERQQHGE
KEEDEGEQPRPFPPRPRQPHQEEHEHQEEHEWHRKEEKHGGKGSEEEQDEREHPRPHQPHQKEEEKHE
WQHKQEKHQGKESEEEEEEDQDEDEGQDKESQSEGESQREPRRHKNKNPFHFN SKRFQTLFKNQYGHVR
VLQRFNKR SQQLQNL RDYRILEFNSKPNTLLSPNHADADYLIVILNGTAILTLVNDDRDSYNLQSGDAL
RVPAGTTYVYVNP DN DENL RMITLAIPV NKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEINK
VLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKPSSRKTISSEDKPFNLRSRDP IYSNKL GKLFEI
TPEKNPQLRDL DVFLSVVDMNEGALFLPHFN SKAIVVLVINEGEANIELVGIKEQQQRQQQEEQPLEVRK
YRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIP SQVQELAFPGSAKD
IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF Y

>BAA74452

VEEEEECEEGQIPRPRQHPERERQQHGEKEEDEGEQPRPFPPRPRQPHQEEHEHQEEHEWHRKEEKH
GGKGSEEEQDEREHPRPHQPHQKEEEKHEWQHKQEKHQGKESEEEEEEDQDEDEEQDKESQSEGESQRE
PRRHKNKNPFHFN SKRFQTLFKNQYGHVRVLQRFNKR SQQLQNL RDYRILEFNSKPNTLLPHHADADYL
IVILNGTAILTLVNDDRDSYNLQSGDALRVPAGTTYVYVNP DN DENL RMITLAIPV NKPGRFESFFLS
TQAQQSYLQGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSR
KTISSEDKPFNLRSRDP IYSNKL GKLFEITPEKNPQLRDL DVFLSVVDMNEGALFLPHFN SKAIVVLV
INEGEANIELVGIKEQQQRQQQEEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNQR
NFLAGSKDNVISQIP SQVQELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF Y

>BAC78524

MMRARFP LLLLGVVFLASVS SFGIAYWEKQNP SHNKL RSCNSEKDSYRNQACHARC NLLKVEEEEECE
EGQIPRPRQHPERERQQHGEKEEDEGEQPRPFPPRPRQPHQEEHEHQEEHEWHRKEEKHGGKGSEEE
QDEREHPRPHQPHQKEEEKHEWQHKQEKHQGKESEEEEEEDQDEDEEQDKESQSEGESQREPRRHKNKN
PFHFN SKRFQTLFKNQYGHVRVLQRFNKR SQQLQNL RDYRILEFNSKPNTLLPHHADADYLIVILNGTA
ILTLVNDDRDSYNLQSGDALRVPAGTTYVYVNP DN DENL RMITLAIPV NKPGRFESFFLSSTQAQQSYL
QGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK
PFNLRSRDP IYSNKL GKLFEITPEKNPQLRDL DVFLSVVDMNEGALFLPHFN SKAIVVLVINEGEANIEL
VGIKEQQQRQQQEEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKD
NVISQIP SQVQELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF Y

>Glyma20g28460.1

MMRVRFPLLVLLGTVFLASVCVSLKVREDENNPFYLRSSNSFQTLFENQNGRIRLLQRFNKRSPQLENLR
DYRIVQFQSKPNTILLPHHADADFLFVLSGRAILTLVNDDRDSYNLHPGDAQRI PAGTTYLVNPHDH
QNLKIIKLAIPV NKPGRYDDFFLSSTQAQQSYLQGFSHNILETSFHSEFEEINRVLFGEEEEQRQQEGVI
VELSKEQIRQLSRRAKSSSRKTISSEDEPFNLRSRNPIYSN NFGKFFEITPEKNPQLRDL DIFLSSVDIN
EGALLPHFN SKAIVILVINEGDANIELVGIKEQQQKQKQEEEPLEVQR YRAELSEDDVFVIPAAYPFVV
NATSNLNLFAGINAENNQRNFLAGEKDNVVRQIERQVQELAFPGSAQDVERLLKKQRESYFVDAQPQQK
EEGSKGRKGPPFSILGAL Y*

>AAB23463

MMRVRFPLLVLLGTVFLASVCVSLKVREDENNPFYFRSSNSFQTLFENQNVIRILLQRFNKRSPQLENLR
DYRIVQFQSKPNTILLPHHADADFLFVLSGRAILTLVNDDRDSYNLHPGDAQRI PAGTTYLVNPHDH
QNLKIIKLAIPV NKPGRYDDFFLSSTQAQQSYLQGFSHNILETSFHSEFEEINRVLFGEEEEQRQQEGVI
VELSKEQIRQLSRRAKSSSRKTISSEDEPFNLRSRNPIYSN NFGKFFEITPEKNPQLRDL DIFLSSVDIN
EGALLPHFN SKAIVILVINEGDANIELVGIKEQQQKQKQEEEPLEVQR YRAELSEDDVFVIPAAYPFVV
NATSNLNLFAGINAENNQRNFLAGEKDNVVRQIERQVQELAFPGSAQDVERLLKKQRESYFVDAQPQQK
EEGSKGRKGPPFSILGAL Y

>Glyma20g28650.1

MMRARFP LLLLGLVFLASVS SFGIAYWEKENPKHNKCLQSCNSERDSYRNQACHARC NLLKVEKEECE
GEIPRPRPRQHPEREPQQGKEKEEDEGEQPRPIPFPRPQPRQEEHEQREEQEWPRKEEKRGEKGSEEE
DEDEDEEQDERQFPFPRPHQKEERKQEEDEDEEQRESESESESELRRHKNKNPFLFGSNRFETL FKNQ
YGRIRVLQRFNQRSPQLQNL RDYRILEFNSKPNTLLPNHADADYLIVILNGTAILSLVNDDRDSYRLQ
SGDALRVP SGTTYVYVNP DN DENL RMITLAIPV NKPGRFESFFLSSTEAQQSYLQGFSRNILEASYDTKF
EINKVLF SREEGQQQGEQRLQESVIVEISKEQIRALSKRAKSSSRKTISSEDKPFNLRSRDP IYSNKL G
KFFEITPEKNPQLRDL DIFLSIVDMNEGALLPHFN SKAIVILVINEGDANIELVGLKEQQQEQQEEQ
LEVRKYRAELSEQDIFVIPAGYPVVVNATSNL NFFAIGINAENNQRNFLAGSQDNVISQIP SQVQELAF
GSAQAVEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILRAF Y*

>Glyma20g28660.1

MMRARFP LLLLGLVFLASVS SFGIAYWEKENPKHNKCLQSCNSERDSYRNQACHARC NLLKVEKEECE
GEIPRPRPRQHPEREPQQGKEKEEDEGEQPRPIPFPRPQPRQEEHEQREEQEWPRKEEKRGEKGSEEE
DEDEDEEQDERQFPFPRPHQKEERKQEEDEDEEQRESESESESELRRHKNKNPFLFGSNRFETL FKNQ
YGRIRVLQRFNQRSPQLQNL RDYRILEFNSKPNTLLPNHADADYLIVILNGTAILSLVNDDRDSYRLQ
SGDALRVP SGTTYVYVNP DN DENL RMITLAIPV NKPGRFESFFLSSTEAQQSYLQGFSRNILEASYDTKF
EINKVLF SREEGQQQGEQRLQESVIVEISKEQIRALSKRAKSSSRKTISSEDKPFNLRSRDP IYSNKL G
KFFEITPEKNPQLRDL DIFLSIVDMNEGALLPHFN SKAIVILVINEGDANIELVGLKEQQQEQQEEQ
LEVRKYRAELSEQDIFVIPAGYPVVVNATSNL NFFAIGINAENNQRNFLAGSQDNVISQIP SQVQELAF
GSAQAVEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILRAF Y*

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

>121281

MMRARFPLLLLGLVFLASVSVSFGIAYWEKENPKHNKCLQSCNSERDSYRNQACHARC�LLKVEKEECE
GEIPRPRPQHPEREPQQGEKEEDEGEQPRPIPPRPPQPRQEEHEQREEQEWPKEEKGEGKSEEE
DEDEDEEQDERQFPFPPRPHQKEERNEEDEDEEQRESESESESELRRHKCNPNFLFGSNRFETLFKNQ
YGRIRVLQRFNQSPQLQNLRDYRILEFNKPNNTLLPNHADADYLIVILNGTAILSLVNNDDRDSYRLQ
SGDALRVPSGTTYVYVNPNDNNENLRLITLAIPVNKPGRFESFLLSSTEAQQSYLQGFSRNILEASYDTKF
EEINKVLFSSREEGQQQGEQRLQESVIVEISKEQIRALSKRAKSSSRKTISSEDKPFNLRSRDIYSNKL
KFFEITPEKNPQLRDLDFLSIVDMNEGALLPHFNKAIIVLVINEGDANIELVGLKEQQQEQQEQEQ
LEVRKYRAELSEQDIFVIPAGYPVVVNATSNLFFAIGINAENNRNFLAGSQDNVISQIPSVQELAF
GSAQAWEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILAFY

>Glyma20g28650.2

MMRARFPLLLLGLVFLASVSVSFGIAYWEKENPKHNKCLQSCNSERDSYRNQACHARC�LLKVEKEECE
GEIPRPRPQHPEREPQQGEKEEDEGEQPRPIPPRPPQPRQEEHEQREEQEWPKEEKGEGKSEEE
DEDEDEEQDERQFPFPPRPHQKEERKQEEDEDEEQRESESESESELRRHKCNPNFLFGSNRFETLFKNQ
YGRIRVLQRFNQSPQLQNLRDYRILEFNKPNNTLLPNHADADYLIVILNGTAILSLVNNDDRDSYRLQ
SGDALRVPSGTTYVYVNPNDNNENLRLITLAIPVNKPGRFESFLLSSTEAQQSYLQGFSRNILEASYDTKF
EEINKVLFSSREEGQQQGEQRLQESVIVEISKEQIRALSKRAKSSSRKTISSEDKPFNLRSRDIYSNKL
KFFEITPEKNPQLRDLDFLSIVDMNEGALLPHFNKAIIVLVINEGDANIELVGLKEQQQEQQEQEQ
LEVRKYRAELSEQDIFVIPAGYPVVVNATSNLFFAIGINAENNRNFLAGI*

>AAA33947

SKRAKSSSRKTISSEDKPFNLGSRDIYSKKLGKFFEITPEKNPQLRDLDFLSIVDMNEGALLPHFN
KAIVLVINEGDANIELVGLKEQQQEQQEQEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSNLFFA
GINAENNRNFLAGSQDNVISQIPSVQELAFPGSAQAWEKLLKNQRESYFVDAQPNEKEEGNKGRKG
PLSSILAFY

>NP_001237316

MMRARFPLLLLGVVFLASVSVSFGIAYWEKQNPŠHNKCLRSCNSEKDSYRNQACHARC�LLKVEEEEECE
EGQIPRPRPQHPERERQQHGEKEEDEGEQPRPFPPRPPRQPRQEGEHEQKEEHEWHRKEEKHGGKGSEEE
QDGREHPRPHQPHQKEEEKHEWQHKKQEKHQGKESEEEEEEDQDEDEEQDKESQESEGSESQREPRRHKN
PFHFNŠKRFQTLFKNQYGHVRLQRFNKRŠQQLQNLRDYRILEFNŠKPNNTLLPHHADADYLIVILNGTA
ILTLVNNDDRDSYNLQSGDALRVAGTTYVYVNPNDNENLRMITLAIPVNKPGRFESFLLSSTQAQQSYL
QGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKRAKSSSRKTISSEDK
PFNLRSRDIYSNKLGLKLFETPEKNPQLRDLDFLSVVDNMNEGALFLPHFNKAIIVLVINEGEANIEL
VGIKEQQQRQQEQEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNRNFLAGSKD
NVISQIPSVQELAFGSAKDİENLİKSQSESYFVDAQPQQKEEGNKGRKGPLSSILAFY

>BAE02726

MMRARFPLLLLGVVFLASVSVSFGIAYWEKQNPŠHNKCLRSCNSEKDSYRNQACHARC�LLKVEEEEECE
EGQIPRPRPQHPERERQQHGEKEEDEGEQPRPFPPRPPRQPHQEEHEQKEEHEWHRKEEKHGGKGSEEE
QDEREHPRPHQPHQKEEEKHEWQHKKQEKHQGKESEEEEEEDQDEDEEQDKESQESEGSESQREPRRHKN
PFHFNŠKRFQTLFKNQYGHVRLQRFNKRŠQQLQNLRDYRILEFNŠKPNNTLLPHHADADYLIVILNGTA
ILTLVNNDDRDSYNLQSGDALRVAGTTYVYVNPNDNENLRMITLAIPVNKPGRFESFLLSSTQAQQSYL
QGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK
PFNLRSRDIYSNKLGLKLFETPEKNPQLRDLDFLSVVDNMNEGALFLPHFNKAIIVLVINEGEANIEL
VGIKEQQQRQQEQEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNRNFLAGSKD
NVISQIPSVQELAFGSAKDİENLİKSQSESYFVDAQPQQKEEGNKGRKGPLSSILAFY

>AAB01374

MMRARFPLLLLGVVFLASVSVSFGIAYWEKQNPŠHNKCLRSCNSEKDSYRNQACHARC�LLKVEEEEECE
EGQIPRPRPQHPERERQQHGEKEEDEGEQPRPFPPRPPRQPHQEEHEQKEEHEWHRKEEKHGGKGSEEE
QDEREHPRPHQPHQKEEEKHEWQHKKQEKHQGKESEEEEEEDQDEDEEQDKESQESEGSESQREPRRHKN
PFHFNŠKRFQTLFKNQYGHVRLQRFNKRŠQQLQNLRDYRILEFNŠKPNNTLLPHHADADYLIVILNGTA
ILTLVNNDDRDSYNLQSGDALRVAGTTYVYVNPNDNENLRMIAGTTYVYVNPNDNENLRMITLAIPVN
PGRFESFLLSSTQAQQSYLQGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIR
ELSKHAKSSSRKTISSEDKPFNLGSRDIYSNKLGLKLFETQRNPQLRDLDFLSVVDNMNEGALFLPHFN
SKAIVLVINEGEANIELVGIKEQQQRQQEQEQPLEVRKYRAELSEQDIFVIPAGYPVMVNATSDLNFFA
FGINAENNRNFLAGSKDNVISQIPSVQELAFPRSAKDİENLİKSQSESYFVDAQPQQKEEGNKGRKG
PLSSILAFY

>Glyma10g39150.1

MMRARFPLLLLGVVFLASVSVSFGIAYWEKQNPŠHNKCLRSCNSEKDSYRNQACHARC�LLKVEEEEECE
EGQIPRPRPQHPERERQQHGEKEEDEGEQPRPFPPRPPRQPHQEEHEQKEEHEWHRKEEKHGGKGSEEE
QDEREHPRPHQPHQKEEEKHEWQHKKQEKHQGKESEEEEEEDQDEDEEQDKESQESEGSESQREPRRHKN
PFHFNŠKRFQTLFKNQYGHVRLQRFNKRŠQQLQNLRDYRILEFNŠKPNNTLLPHHADADYLIVILNGTA
ILTLVNNDDRDSYNLQSGDALRVAGTTYVYVNPNDNENLRMITLAIPVNKPGRFESFLLSSTQAQQSYL
QGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK
PFNLRSRDIYSNKLGLKLFETPEKNPQLRDLDFLSVVDNMNEGALFLPHFNKAIIVLVINEGEANIEL
VGIKEQQQRQQEQEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNRNFLAGSKD
NVISQIPSVQELAFGSAKDİENLİKSQSESYFVDAQPQQKEEGNKGRKGPLSSILAFY*

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

Gly m 5 (beta-conglycinin) Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

		1		50
Gly m 5 ABH09130	(1)	-----MFGIVYWEKQNP	SHNKCLRSCNSEKDSY	
Gly m 5 BAA74452	(1)	-----		
Gly m 5 BAC78524	(1)	MMRARFPLLLLG-VVFLASVSVSFGIAYWEKQNP	SHNKCLRSCNSEKDSY	
Gly m 5 Glyma20g28460.1	(1)	MMRVRFPLLVLVLLGTVFLASVCVS	-----	
Gly m 5 AAB23463	(1)	MMRVRFPLLVLVLLGTVFLASVCVS	-----	
Gly m 5 Glyma20g28650.1	(1)	MMRARFPLLLLG-LVFLASVSVSFGIAYWEKENPK	HNKCLQSCNSEKDSY	
Gly m 5 Glyma20g28660.1	(1)	MMRARFPLLLLG-LVFLASVSVSFGIAYWEKENPK	HNKCLQSCNSEKDSY	
Gly m 5 121281	(1)	MMRARFPLLLLG-LVFLASVSVSFGIAYWEKENPK	HNKCLQSCNSEKDSY	
Gly m 5 Glyma20g28650.2	(1)	MMRARFPLLLLG-LVFLASVSVSFGIAYWEKENPK	HNKCLQSCNSEKDSY	
Gly m 5 AAA33947	(1)	-----		
Gly m 5 NP_001237316	(1)	MMRARFPLLLLG-VVFLASVSVSFGIAYWEKQNP	SHNKCLRSCNSEKDSY	
Gly m 5 BAE02726	(1)	MMRARFPLLLLG-VVFLASVSVSFGIAYWEKQNP	SHNKCLRSCNSEKDSY	
Gly m 5 AAB01374	(1)	MMRARFPLLLLG-VVFLASVSVSFGIAYWEKQNP	SHNKCLRSCNSEKDSY	
Gly m 5 Glyma10g39150.1	(1)	MMRARFPLLLLG-VVFLASVSVSFGIAYWEKQNP	SHNKCLRSCNSEKDSY	
Consensus	(1)	MMRARFPLLLLG VFLASVSVSFGIAYWEK NP HNKCL SCNSE DSY		
		51		100
Gly m 5 ABH09130	(29)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 BAA74452	(1)	-----VEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 BAC78524	(50)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 Glyma20g28460.1	(24)	-----LKVR	ED	
Gly m 5 AAB23463	(24)	-----LKVR	ED	
Gly m 5 Glyma20g28650.1	(50)	RNQACHARNLLKVEKEE-CEEGEIPRPRPRP	QHPEREQQPGEKEEDE	
Gly m 5 Glyma20g28660.1	(50)	RNQACHARNLLKVEKEE-CEEGEIPRPRPRP	QHPEREQQPGEKEEDE	
Gly m 5 121281	(50)	RNQACHARNLLKVEKEE-CEEGEIPRPRPRP	QHPEREQQPGEKEEDE	
Gly m 5 Glyma20g28650.2	(50)	RNQACHARNLLKVEKEE-CEEGEIPRPRPRP	QHPEREQQPGEKEEDE	
Gly m 5 AAA33947	(1)	-----		
Gly m 5 NP_001237316	(50)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 BAE02726	(50)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 AAB01374	(50)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Gly m 5 Glyma10g39150.1	(50)	RNQACHARNLLKVEEEEECEEGQIPRPRP	QHPERERQQHGEKEEDEG	
Consensus	(51)	RNQACHARNLLKVEEEEECEEGQIPRPRP QHPERERQQHGEKEEDEG		
		101		150
Gly m 5 ABH09130	(77)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 BAA74452	(36)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 BAC78524	(98)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 Glyma20g28460.1	(30)	E-----		
Gly m 5 AAB23463	(30)	E-----		
Gly m 5 Glyma20g28650.1	(99)	EQPRIPFPRP-QPRQEEEHEQREEQEWPRKEEKRGEKGSEEEDEDEDEE		
Gly m 5 Glyma20g28660.1	(99)	EQPRIPFPRP-QPRQEEEHEQREEQEWPRKEEKRGEKGSEEEDEDEDEE		
Gly m 5 121281	(99)	EQPRIPFPRP-QPRQEEEHEQREEQEWPRKEEKRGEKGSEEEDEDEDEE		
Gly m 5 Glyma20g28650.2	(99)	EQPRIPFPRP-QPRQEEEHEQREEQEWPRKEEKRGEKGSEEEDEDEDEE		
Gly m 5 AAA33947	(1)	-----		
Gly m 5 NP_001237316	(98)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 BAE02726	(98)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 AAB01374	(98)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Gly m 5 Glyma10g39150.1	(98)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE	-----E	
Consensus	(101)	EQPRFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEE		E

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

		151	200
Gly m 5 ABH09130	(120)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEGQDKE	
Gly m 5 BAA74452	(79)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Gly m 5 BAC78524	(141)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Gly m 5 Glyma20g28460.1	(31)	-----N-----	
Gly m 5 AAB23463	(31)	-----N-----	
Gly m 5 Glyma20g28650.1	(148)	QDERQF PFPRP PHQKEER--KQEE-----DEDEEQQRE	
Gly m 5 Glyma20g28660.1	(148)	QDERQF PFPRP PHQKEER--KQEE-----DEDEEQQRE	
Gly m 5 121281	(148)	QDERQF PFPRP PHQKEER--NEEE-----DEDEEQQRE	
Gly m 5 Glyma20g28650.2	(148)	QDERQF PFPRP PHQKEER--KQEE-----DEDEEQQRE	
Gly m 5 AAA33947	(1)	-----	
Gly m 5 NP_001237316	(141)	QDGREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Gly m 5 BAE02726	(141)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Gly m 5 AAB01374	(141)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Gly m 5 Glyma10g39150.1	(141)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
Consensus	(151)	QDEREHPRPHQPHQKEEEKHEWQHKEKHQKGKSEEEEEEDQDEDEEQDKE	
		201	250
Gly m 5 ABH09130	(170)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 BAA74452	(129)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 BAC78524	(191)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 Glyma20g28460.1	(32)	-----NPFYLRSSNSFQTLFENQNGRI RLQRFNKR	
Gly m 5 AAB23463	(32)	-----NPFYFRSSNSFQTLFENQNVRI RLQRFNKR	
Gly m 5 Glyma20g28650.1	(179)	SEESE----DS ELRRHKNKNPFLFGS-NRFE TLFKNQYGRI RVLQRFNQR	
Gly m 5 Glyma20g28660.1	(179)	SEESE----DS ELRRHKNKNPFLFGS-NRFE TLFKNQYGRI RVLQRFNQR	
Gly m 5 121281	(179)	SEESE----DS ELRRHKNKNPFLFGS-NRFE TLFKNQYGRI RVLQRFNQR	
Gly m 5 Glyma20g28650.2	(179)	SEESE----DS ELRRHKNKNPFLFGS-NRFE TLFKNQYGRI RVLQRFNQR	
Gly m 5 AAA33947	(1)	-----	
Gly m 5 NP_001237316	(191)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 BAE02726	(191)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 AAB01374	(191)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Gly m 5 Glyma10g39150.1	(191)	SQSEGESSESQREPRRHKNKNPFHFNS-KRFQTLFKNQYGHVRLQRFNKR	
Consensus	(201)	SQSEGESSESQREPRRHKNKNPFHFNS KRFQTLFKNQYGHVRLQRFNKR	
		251	300
Gly m 5 ABH09130	(219)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 BAA74452	(178)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 BAC78524	(240)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 Glyma20g28460.1	(63)	SPQLENLRDYRIVQFQSKPNTI LLPHHADADFLFVLSGRAILTLVNND	
Gly m 5 AAB23463	(63)	SPQLENLRDYRIVQFQSKPNTI LLPHHADADFLFVLSGRAILTLVNND	
Gly m 5 Glyma20g28650.1	(224)	SPQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILSLVNND	
Gly m 5 Glyma20g28660.1	(224)	SPQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILSLVNND	
Gly m 5 121281	(224)	SPQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILSLVNND	
Gly m 5 Glyma20g28650.2	(224)	SPQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILSLVNND	
Gly m 5 AAA33947	(1)	-----	
Gly m 5 NP_001237316	(240)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 BAE02726	(240)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 AAB01374	(240)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Gly m 5 Glyma10g39150.1	(240)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	
Consensus	(251)	SQQLQNLRDYRILEFNSKPNNTLLSPNHADADYLIVILNGTAILTLVNND	

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

		301		350
Gly m 5 ABH09130	(269)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Gly m 5 BAA74452	(228)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Gly m 5 BAC78524	(290)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Gly m 5 Glyma20g28460.1	(113)	RDSYNLHPGDAQRI PAGTTYLVN	-----	PHDHQNL
Gly m 5AAB23463	(113)	RDSYNLHPGDAQRI PAGTTYLVN	-----	PHDHQNL
Gly m 5 Glyma20g28650.1	(274)	RDSYRLQSGDALRVPSGTTYVNVN	-----	PDNNENL
Gly m 5 Glyma20g28660.1	(274)	RDSYRLQSGDALRVPSGTTYVNVN	-----	PDNNENL
Gly m 5 121281	(274)	RDSYRLQSGDALRVPSGTTYVNVN	-----	PDNNENL
Gly m 5 Glyma20g28650.2	(274)	RDSYRLQSGDALRVPSGTTYVNVN	-----	PDNNENL
Gly m 5 AAA33947	(1)	-----	-----	-----
Gly m 5 NP_001237316	(290)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Gly m 5 BAE02726	(290)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Gly m 5 AAB01374	(290)	RDSYNLQSGDALRVPAGTTFYVNVN	PDNDENLRMIAGTTFYVNVN	PDNDENL
Gly m 5 Glyma10g39150.1	(290)	RDSYNLQSGDALRVPAGTTYVNVN	-----	PDNDENL
Consensus	(301)	RDSYNLQSGDALRVPAGTTYVNVN		PDNDENL
		351		400
Gly m 5 ABH09130	(300)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 BAA74452	(259)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 BAC78524	(321)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 Glyma20g28460.1	(144)	KI IKLAIPVNKPGRYDD FFLSSTQAQQSYLQGFSHNILETSFHSEFEEIN		
Gly m 5AAB23463	(144)	KI IKLAIPVNKPGRYDD FFLSSTQAQQSYLQGFSHNILETSFHSEFEEIN		
Gly m 5 Glyma20g28650.1	(305)	RLITLAIPVNKPGRFESFFLSSTE AQQSYLQGFSRNILEASYDTKFEEIN		
Gly m 5 Glyma20g28660.1	(305)	RLITLAIPVNKPGRFESFFLSSTE AQQSYLQGFSRNILEASYDTKFEEIN		
Gly m 5 121281	(305)	RLITLAIPVNKPGRFESFFLSSTE AQQSYLQGFSRNILEASYDTKFEEIN		
Gly m 5 Glyma20g28650.2	(305)	RLITLAIPVNKPGRFESFFLSSTE AQQSYLQGFSRNILEASYDTKFEEIN		
Gly m 5 AAA33947	(1)	-----	-----	-----
Gly m 5 NP_001237316	(321)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 BAE02726	(321)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 AAB01374	(340)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Gly m 5 Glyma10g39150.1	(321)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
Consensus	(351)	RMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEIN		
		401		450
Gly m 5 ABH09130	(350)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKPSSRKTISSEDK		
Gly m 5 BAA74452	(309)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK		
Gly m 5 BAC78524	(371)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK		
Gly m 5 Glyma20g28460.1	(194)	RVLFGEEE-----EQRQQEGVIVELSKEQIRQLSRRAKSSSRKTISSEDE		
Gly m 5AAB23463	(194)	RVLFGEEE-----EQRQQEGVIVELSKEQIRQLSRRAKSSSRKTISSEDE		
Gly m 5 Glyma20g28650.1	(355)	KVLFSREEGQQQGEQRLQESVIVEISKEQIRALS KRAKSSSRKTISSEDK		
Gly m 5 Glyma20g28660.1	(355)	KVLFSREEGQQQGEQRLQESVIVEISKEQIRALS KRAKSSSRKTISSEDK		
Gly m 5 121281	(355)	KVLFSREEGQQQGEQRLQESVIVEISKEQIRALS KRAKSSSRKTISSEDK		
Gly m 5 Glyma20g28650.2	(355)	KVLFSREEGQQQGEQRLQESVIVEISKEQIRALS KRAKSSSRKTISSEDK		
Gly m 5 AAA33947	(1)	-----SKRAKSSSRKTISSEDK		
Gly m 5 NP_001237316	(371)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKRAKSSSRKTISSEDK		
Gly m 5 BAE02726	(371)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK		
Gly m 5 AAB01374	(390)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK		
Gly m 5 Glyma10g39150.1	(371)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK		
Consensus	(401)	KVLFGREEGQQQGEERLQESVIVEISKKQIRELSKRAKSSSRKTISSEDK		

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

		451		500
Gly m 5 ABH09130	(400)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 BAA74452	(359)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 BAC78524	(421)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 Glyma20g28460.1	(239)	PFNLRSRNPIYSNNFGKF FEITPEKNPQLRDLDIFLSSVDINEGALLLPH		
Gly m 5 AAB23463	(239)	PFNLRSRNPIYSNNFGKF FEITPEKNPQLRDLDIFLSSVDINEGALLLPH		
Gly m 5 Glyma20g28650.1	(405)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDIFLSIVDMNEGALLLPH		
Gly m 5 Glyma20g28660.1	(405)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDIFLSIVDMNEGALLLPH		
Gly m 5 121281	(405)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDIFLSIVDMNEGALLLPH		
Gly m 5 Glyma20g28650.2	(405)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDIFLSIVDMNEGALLLPH		
Gly m 5 AAA33947	(18)	PFNLGSRDPIYSKLGKFEITPEKNPQLRDLDIFLSIVDMNEGALLLPH		
Gly m 5 NP_001237316	(421)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 BAE02726	(421)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 AAB01374	(440)	PFNLGSRDPIYSNKLGLFEITQ-RNPQLRDLDVFLSVVDMNEGALFLPH		
Gly m 5 Glyma10g39150.1	(421)	PFNLRSRDPIYSNKLGLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPH		
Consensus	(451)	PFNLRSRDPIYSNKLGL FEITPEKNPQLRDLD FLSVVDMNEGAL LPH		
		501		550
Gly m 5 ABH09130	(450)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 BAA74452	(409)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 BAC78524	(471)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 Glyma20g28460.1	(289)	FNSKAIVILVINEGDANIELVG IKEQQQKQKQEEPLEVQRYRAELSEDD		
Gly m 5 AAB23463	(289)	FNSKAIVILVINEGDANIELVG IKEQQQKQKQEEPLEVQRYRAELSEDD		
Gly m 5 Glyma20g28650.1	(455)	FNSKAIVILVINEGDANIELVGLKEQQQEQQQEEQPLEVRKYRAELSEQD		
Gly m 5 Glyma20g28660.1	(455)	FNSKAIVILVINEGDANIELVGLKEQQQEQQQEEQPLEVRKYRAELSEQD		
Gly m 5 121281	(455)	FNSKAIVILVINEGDANIELVGLKEQQQEQQQEEQPLEVRKYRAELSEQD		
Gly m 5 Glyma20g28650.2	(455)	FNSKAIVILVINEGDANIELVGLKEQQQEQQQEEQPLEVRKYRAELSEQD		
Gly m 5 AAA33947	(68)	FNSKAIVILVINEGDANIELVGLKEQQQEQQQEEQPLEVRKYRAELSEQD		
Gly m 5 NP_001237316	(471)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 BAE02726	(471)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 AAB01374	(489)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Gly m 5 Glyma10g39150.1	(471)	FNSKAIVVLVINEGEANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
Consensus	(501)	FNSKAIV LVINEG ANIELVG IKEQQQRQQQEEQPLEVRKYRAELSEQD		
		551		600
Gly m 5 ABH09130	(500)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 BAA74452	(459)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 BAC78524	(521)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 Glyma20g28460.1	(339)	VFVIPAAYPFVVNATSNLNF LAFGINAENNQRNFLAGSKDNVVRQIERQV		
Gly m 5 AAB23463	(339)	VFVIPAAYPFVVNATSNLNF LAFGINAENNQRNFLAGSKDNVVRQIERQV		
Gly m 5 Glyma20g28650.1	(505)	IFVIPAGYPVVVNATSNLNF FAIGINAENNQRNFLAGSDNVISQIPSQV		
Gly m 5 Glyma20g28660.1	(505)	IFVIPAGYPVVVNATSNLNF FAIGINAENNQRNFLAGSDNVISQIPSQV		
Gly m 5 121281	(505)	IFVIPAGYPVVVNATSNLNF FAIGINAENNQRNFLAGSDNVISQIPSQV		
Gly m 5 Glyma20g28650.2	(505)	IFVIPAGYPVVVNATSNLNF FAIGINAENNQRNFLAGI-----		
Gly m 5 AAA33947	(118)	IFVIPAGYPVVVNATSNLNF FAIGINAENNQRNFLAGSDNVISQIPSQV		
Gly m 5 NP_001237316	(521)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 BAE02726	(521)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 AAB01374	(539)	IFVIPAGYPVMVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Gly m 5 Glyma10g39150.1	(521)	IFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
Consensus	(551)	IFVIPAGYPVVVNATSD LNFFAFGINAENNQRNFLAGSKDNVISQIPSQV		
		601		650

Supporting Information S4 (Continued): Gly m 5 Consensus Sequence Generation

Gly m 5 ABH09130	(550)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 BAA74452	(509)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 BAC78524	(571)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 Glyma20g28460.1	(389)	QELAFPGSAQDVERLLKKQRESYFVDAQPQQKEEGSKGRKGPFPSILGAL
Gly m 5 AAB23463	(389)	QELAFPGSAQDVERLLKKQRESYFVDAQPQQKEEGSKGRKGPFPSILGAL
Gly m 5 Glyma20g28650.1	(555)	QELAFPGSAQAVEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILRAF
Gly m 5 Glyma20g28660.1	(555)	QELAFPGSAQAVEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILRAF
Gly m 5 121281	(555)	QELAFPGSAQAVEKLLKNQRESYFVDAQPKKKEEGNKGRKGPLSSILRAF
Gly m 5 Glyma20g28650.2	(543)	-----
Gly m 5 AAA33947	(168)	QELAFPGSAQAVEKLLKNQRESYFVDAQPNEKEEGNKGRKGPLSSILRAF
Gly m 5 NP_001237316	(571)	QELAF LGS AKDIENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 BAE02726	(571)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 AAB01374	(589)	QELAFPRSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Gly m 5 Glyma10g39150.1	(571)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
Consensus	(601)	QELAFPGSAKD IENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAF
	651	
Gly m 5 ABH09130	(600)	Y
Gly m 5 BAA74452	(559)	Y
Gly m 5 BAC78524	(621)	Y
Gly m 5 Glyma20g28460.1	(439)	Y
Gly m 5 AAB23463	(439)	Y
Gly m 5 Glyma20g28650.1	(605)	Y
Gly m 5 Glyma20g28660.1	(605)	Y
Gly m 5 121281	(605)	Y
Gly m 5 Glyma20g28650.2	(543)	-
Gly m 5 AAA33947	(218)	Y
Gly m 5 NP_001237316	(621)	Y
Gly m 5 BAE02726	(621)	Y
Gly m 5 AAB01374	(639)	Y
Gly m 5 Glyma10g39150.1	(621)	Y
Consensus	(651)	Y

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 5 121281 605aa
MMRARFP L L L L L G L V F L A S V S V S F G I A Y W E K E N P K H N K C L Q S C N S E R D S Y R N Q A C H A R C N L L K V E K E E C E E
G E I P R P R P R P Q H P E R E P Q Q G E K E E D E D E Q P R P I P F P R P Q P R Q E E E H E Q R E E Q E W P R K E E K R G E K G S E E E
D E D E D E E Q D E R Q F P F P R P P H Q K E E R N E E D E D E E Q Q R E S E E S E D S E L R R H K N K N P F L F G S N R F E T L F K N Q
Y G R I R V L Q R F N Q R S P Q L Q N L R D Y R I L E F N S K P N T L L P N H A D A D Y L I V I L N G T A I L S L V N N D D R D S Y R L Q
S G D A L R V P S G T T Y Y V V N P D N N E N L R L I T A I P V N K P G R F E S F L S S T E A Q Q S Y L Q G F S R N I L E A S Y D T K F
E E I N K V L F S R E E G Q Q Q G E Q R L Q E S V I E I S K E Q I R A L S K R A K S S R K T I S S E D K P F N L R S R D P I Y S N K L G
K F F E I T P E K N P Q L R D L D I F L S I V D M N E G A L L P H F N S K A I V I L V I N E G D A N I E L V G L K E Q Q Q E Q Q Q E E Q P
L E V R K Y R A E L S E Q D I F V I P A G Y P V V V N A T S N L N F F A I G I N A E N N Q R N F L A G S Q D N V I S Q I P S Q V Q E L A F P
G S A Q A V E K L L K N Q R E S Y F V D A Q P K K K E E G N K G R K G P L S S I L R A F Y

Supporting Information S5: Gly m 6 (Glycinin G1 Subunit) Consensus Sequence Generation

Gly m 6 G1 subunit Protein Sequences: (highlighting colors indicate shared peptide sequences among available proteins)

>Gly m 6 Glycinin G1 495aa

MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGGLIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNQGPQEIYIQQKGKIFGMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGD
LIAVPTGVAVWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQGGHQSQKKGKHQ
QEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQRPQEEE
EEEEDEKQPCKGKDKHCQRPRGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP
ALSWLRLSAEFGSLRKNAMFVPHYNLANSIIYALNGRALIQVVCNNGERVFDGELQEGRVLIVPQNFVV
AARSQSDNFEYVSFKTNDTPMIGTLAGANSLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQ
KRAVA

>Gly m 6 CAA33215.1 subunit G1 495aa

MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGGLIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNQGPQEIYIQQKGKIFGMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGD
LIAVPTGVAVWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQGGHQSQKKGKHQ
QEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQRPQEEE
EEEEDEKQPCKGKDKHCQRPRGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP
ALSWLRLSAEFGSLRKNAMFVPHYNLANSIIYALNGRALIQVVCNNGERVFDGELQEGRVLIVPQNFVV
AARSQSDNFEYVSFKTNDTPMIGTLAGANSLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQ
KRAVA

>Gly m 6 CAA26723.1 495aa

MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPGNRIESEGGLIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNQGPQEIYIQQKGKIFGMIYPGCSSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGD
LIAVPTGVAVWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQGGHQSQKKGKHQ
QEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQRPQEEE
EEEEDEKQPCKGKDKHCQRPRGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP
ALSWLRLSAGFGLRKNAMFVPHYNLANSIIYALNGRALIQVVCNNGERVFDGELQEGRVLIVPQNFVV
AARSQSDNFEYVSFKTNDTPMIGTLAGANSLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQ
KRAVA

Gly m 6 G1 subunit Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
Gly m 6	Glycinin G1	(1)	MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGG	
Gly m 6 CAA33215.1 subunit G1		(1)	MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGG	
Gly m 6 CAA26723.1		(1)	MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGG	
	Consensus	(1)	MAKLVSFLCFLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGG	
			51	100
Gly m 6	Glycinin G1	(51)	LIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQGPQEIYIQQKGKIF	
Gly m 6 CAA33215.1 subunit G1		(51)	LIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQGPQEIYIQQKGKIF	
Gly m 6 CAA26723.1		(51)	LIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQGPQEIYIQQKGKIF	
	Consensus	(51)	LIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQGPQEIYIQQKGKIF	
			101	150
Gly m 6	Glycinin G1	(101)	GMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGDLIAVPTGVAV	
Gly m 6 CAA33215.1 subunit G1		(101)	GMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGDLIAVPTGVAV	
Gly m 6 CAA26723.1		(101)	GMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGDLIAVPTGVAV	
	Consensus	(101)	GMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGDLIAVPTGVAV	
			151	200
Gly m 6	Glycinin G1	(151)	WMYNNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQG	
Gly m 6 CAA33215.1 subunit G1		(151)	WMYNNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQG	
Gly m 6 CAA26723.1		(151)	WMYNNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQG	
	Consensus	(151)	WMYNNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQG	
			201	250
Gly m 6	Glycinin G1	(201)	GHQSQKGKHQEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGE	
Gly m 6 CAA33215.1 subunit G1		(201)	GHQSQKGKHQEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGE	
Gly m 6 CAA26723.1		(201)	GHQSQKGKHQEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGE	
	Consensus	(201)	GHQSQKGKHQEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGE	
			251	300
Gly m 6	Glycinin G1	(251)	DKGAIIVTVKGGLSVIKPPTDEQQQRPQEEEEEEDEKQPCKGKDKHCQRP	
Gly m 6 CAA33215.1 subunit G1		(251)	DKGAIIVTVKGGLSVIKPPTDEQQQRPQEEEEEEDEKQPCKGKDKHCQRP	
Gly m 6 CAA26723.1		(251)	DKGAIIVTVKGGLSVIKPPTDEQQQRPQEEEEEEDEKQPCKGKDKHCQRP	
	Consensus	(251)	DKGAIIVTVKGGLSVIKPPTDEQQQRPQEEEEEEDEKQPCKGKDKHCQRP	
			301	350
Gly m 6	Glycinin G1	(301)	RGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP	
Gly m 6 CAA33215.1 subunit G1		(301)	RGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP	
Gly m 6 CAA26723.1		(301)	RGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP	
	Consensus	(301)	RGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP	

Supporting Information S4 (Continued): Gly m 6 (Glycinin G1 Subunit) Consensus Sequence Generation

			351		400
Gly m 6	Glycinin G1	(351)	ALSWLRLSAEFGSLRKNAMEFVPHYNLNANSIIYALNGRALIQVVNCNGER		
Gly m 6 CAA33215.1	subunit G1	(351)	ALSWLRLSAEFGSLRKNAMEFVPHYNLNANSIIYALNGRALIQVVNCNGER		
Gly m 6 CAA26723.1		(351)	ALSWLRLSAGFGSLRKNAMEFVPHYNLNANSIIYALNGRALIQVVNCNGER		
	Consensus	(351)	ALSWLRLSAEFGSLRKNAMEFVPHYNLNANSIIYALNGRALIQVVNCNGER		
			401		450
Gly m 6	Glycinin G1	(401)	VFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFKTNDTPMIGTLAGANS		
Gly m 6 CAA33215.1	subunit G1	(401)	VFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFKTNDTPMIGTLAGANS		
Gly m 6 CAA26723.1		(401)	VFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFKTNDTPMIGTLAGANS		
	Consensus	(401)	VFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFKTNDTPMIGTLAGANS		
			451		495
Gly m 6	Glycinin G1	(451)	LLNALPEEVIQHTFNLKSQQARQIKNNNPFFKFLVPPQESQKRAVA		
Gly m 6 CAA33215.1	subunit G1	(451)	LLNALPEEVIQHTFNLKSQQARQIKNNNPFFKFLVPPQESQKRAVA		
Gly m 6 CAA26723.1		(451)	LLNALPEEVIQHTFNLKSQQARQIKNNNPFFKFLVPPQESQKRAVA		
	Consensus	(451)	LLNALPEEVIQHTFNLKSQQARQIKNNNPFFKFLVPPQESQKRAVA		

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 6 Glycinin G1 495aa
MAKLVSFLCFLLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGGLIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNGPQEIIYIQQKGIFGMIYPGCPSTFEPPQQPQQRGQSSRPQDRHQKIYNFREGD
LIAVPTGVAVWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQGGHQSQKKGKHQ
QEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQRPQEEE
EEEEDEKPQCKGKDKHCQRPRGSQSKSRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP
ALSWLRLSAEFGSLRKNAMEFVPHYNLNANSIIYALNGRALIQVVNCNGERVFDGELQEGRVLIVPQNFVV
AARSQSDNFEYVSFKTNDTPMIGTLAGANSLLNALPEEVIQHTFNLKSQQARQIKNNNPFFKFLVPPQESQ
KRAVA

Supporting Information S6: Gly m 6 (Glycinin G2 Subunit) Consensus Sequence Generation

Gly m 6 G2 subunit Protein Sequences: (highlighting colors indicate shared peptide sequences among available proteins)

>Gly m 6 Glycinin G2 485aa
 MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT
 LNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIA
 VPTGVAWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEE
 ENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSGAIVTVKGGLRVTAPAMRKPPQEEEDDDDEEE
 QPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ
 YGSLRKNAMFVPHYTLNANSIYALNGRALVQVNCNGERVFDGELQEGGLVLPQNFAVAAKSQSDNFE
 YVSFKTNDRPSIGNLAGANSLNALPEEVIQHTFNLKSQARQVKNNNPFSLVPPQESQRRRAVA

>Gly m 6 CAA33216.1 subunit G2 485aa
 MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT
 LNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIA
 VPTGVAWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEE
 ENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSGAIVTVKGGLRVTAPAMRKPPQEEEDDDDEEE
 QPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ
 YGSLRKNAMFVPHYTLNANSIYALNGRALVQVNCNGERVFDGELQEGGLVLPQNFAVAAKSQSDNFE
 YVSFKTNDRPSIGNLAGANSLNALPEEVIQHTFNLKSQARQVKNNNPFSLVPPQESQRRRAVA

>Gly m 6 CAA26575.1 485aa
 MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT
 LNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIA
 VPTGVAWWMYNNEDTPVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEE
 ENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSGAIVTVKGGLRVTAPAMRKPPQEEEDDDDEEE
 QPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ
 YGSLRKNAMFVPHYTLNANSIYALNGRALVQVNCNGERVFDGELQEGGLVLPQNFAVAAKSQSDNFE
 YVSFKTNDRPSIGNLAGANSLNALPEEVIQHTFNLKSQARQVKNNNPFSLVPPQESQRRRAVA

Gly m 6 G2 subunit Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
Gly m 6	Glycinin G2	(1)	MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT	
Gly m 6	CAA33216.1 subunit G2	(1)	MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT	
	Gly m 6	(1)	MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT	
	Consensus	(1)	MAKLVLSLCFLLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT	
			51	100
Gly m 6	Glycinin G2	(51)	TWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
Gly m 6	CAA33216.1 subunit G2	(51)	TWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
	Gly m 6	(51)	TWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
	Consensus	(51)	TWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNQPQEIYIQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
			101	150
Gly m 6	Glycinin G2	(101)	FPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
Gly m 6	CAA33216.1 subunit G2	(101)	FPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
	Gly m 6	(101)	FPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
	Consensus	(101)	FPGCPSTYQEPQESQQRGRSQRPPDRHQKVHRFREGDLIAVPTGVAWWMY	
			151	200
Gly m 6	Glycinin G2	(151)	NNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
Gly m 6	CAA33216.1 subunit G2	(151)	NNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
	Gly m 6	(151)	NNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
	Consensus	(151)	NNEDTPVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSSQSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
			201	250
Gly m 6	Glycinin G2	(201)	QSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
Gly m 6	CAA33216.1 subunit G2	(201)	QSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
	Gly m 6	(201)	QSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
	Consensus	(201)	QSQKQKQEEENEGSNILSGFAPEFLKEAFGVNMQIVRNQLQGENEEDSG	
			251	300
Gly m 6	Glycinin G2	(251)	AIVTVKGGLRVTAPAMRKPPQEEEDDDDEEEQPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
Gly m 6	CAA33216.1 subunit G2	(251)	AIVTVKGGLRVTAPAMRKPPQEEEDDDDEEEQPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
	Gly m 6	(251)	AIVTVKGGLRVTAPAMRKPPQEEEDDDDEEEQPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
	Consensus	(251)	AIVTVKGGLRVTAPAMRKPPQEEEDDDDEEEQPQCVDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
			301	350
Gly m 6	Glycinin G2	(301)	GIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
Gly m 6	CAA33216.1 subunit G2	(301)	GIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
	Gly m 6	(301)	GIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
	Consensus	(301)	GIDETICTMRLRQNIQNSSPDIYNPQAGSITTATSLDFPALWLLKLSAQ	
			351	400
Gly m 6	Glycinin G2	(351)	YGSLRKNAMFVPHYTLNANSIYALNGRALVQVNCNGERVFDGELQEGGLVLPQNFAVAAKSQSDNFEYVSFKTNDRPSIGNLAGANSLNALPEEVIQHTFNLKSQARQVKNNNPFSLVPPQESQRRRAVA	

Supporting Information S6: Gly m 6 (Glycinin G2 Subunit) Consensus Sequence Generation

Gly m 6	CAA33216.1	subunit G2	(351)	YGSLRKNAMEFVPHYTLNANSIIYALNGRALVQVVNCNGERVFDGELQEGG
	Gly m 6	CAA26575.1	(351)	YGSLRKNAMEFVPHYTLNANSIIYALNGRALVQVVNCNGERVFDGELQEGG
		Consensus	(351)	YGSLRKNAMEFVPHYTLNANSIIYALNGRALVQVVNCNGERVFDGELQEGG
			401	450
	Gly m 6	Glycinin G2	(401)	VLIVPQNFAVAAKSQSDNFEYVSFKTNDRPSIGNLAGANSLLNALPEEVI
Gly m 6	CAA33216.1	subunit G2	(401)	VLIVPQNFAVAAKSQSDNFEYVSFKTNDRPSIGNLAGANSLLNALPEEVI
	Gly m 6	CAA26575.1	(401)	VLIVPQNFAVAAKSQSDNFEYVSFKTNDRPSIGNLAGANSLLNALPEEVI
		Consensus	(401)	VLIVPQNFAVAAKSQSDNFEYVSFKTNDRPSIGNLAGANSLLNALPEEVI
			451	485
	Gly m 6	Glycinin G2	(451)	QHTFNLKSQQARQVKNNNPFSFLVPPQESQRRAVA
Gly m 6	CAA33216.1	subunit G2	(451)	QHTFNLKSQQARQVKNNNPFSFLVPPQESQRRAVA
	Gly m 6	CAA26575.1	(451)	QHTFNLKSQQARQVKNNNPFSFLVPPQESQRRAVA
		Consensus	(451)	QHTFNLKSQQARQVKNNNPFSFLVPPQESQRRAVA

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 6 Glycinin G2 485aa
MAKLVLSCFLFSGCFALREQAQQNECQIQKLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCT
LNRNALRRPSYTNQPQEIIYQQNGIFGMIFPGCPSTYQEPQESQQRGRSQRPQDRHQKVHRFREGDLIA
VPTGVAVWWMYNNEDTPVVAVSIIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQQQQGGSQSQKGKQQEE
ENEGSNILSGFAPEFLKEAFGVNMQIVRNLQGENEEEDSGAIVTVKGGLRVTAPAMRKPPQEEEDDDDEEE
QPQCVELDKGCQRQSKRSRNGIDETICTMRLRQNIQNSSPDYNPQAGSITTATSLDFPALWLLKLSAQ
YGSLRKNAMEFVPHYTLNANSIIYALNGRALVQVVNCNGERVFDGELQEGGVLIIVPQNFAVAAKSQSDNFE
YVSFKTNDRPSIGNLAGANSLLNALPEEVIQHTFNLKSQQARQVKNNNPFSFLVPPQESQRRAVA

Supporting Information S7: Gly m 6 (Glycinin G3 Subunit) Consensus Sequence Generation

Gly m 6 G3 subunit Protein Sequences:

>Gly m 6 Glycinin G3 481aa

MAKLVLSLCFLLFSGCCFAFSFREQPQNECQIQLRNALKPDNRIESEGFIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNAPQEIYIQQSGIFGMIFPGCPSTFEPPQKQGSSRPQDRHQKIYHFREGDLIA
VPTGFAYWMYNNEDTPVVAVSLIDTNSFQNLQDQMPPRRFYLAGNQEQEFLQYQPQKQQGGTQSQKGKRQQ
EEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLGGENEEEEKGAIVTVKGGLSVISPTTEEQQQRPEEEEEK
PDCDEKDKHCQSQSRNGIDETICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSL
RKNAMFVPHYNLNANSIIYALNGRALVQVVCNNGERVFDGELQEGQVLIVPQNFAVAARSQSDNFEYVSF
KTNDRPSIGNLAGANSLLNALPEEVIQQTENLRRQARQVKNNNPFSFLVPPKESQRRVVA

>Gly m 6 CAA33217.1 subunit G3 481aa

MAKLVLSLCFLLFSGCCFAFSFREQPQNECQIQLRNALKPDNRIESEGFIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNAPQEIYIQQSGIFGMIFPGCPSTFEPPQKQGSSRPQDRHQKIYHFREGDLIA
VPTGFAYWMYNNEDTPVVAVSLIDTNSFQNLQDQMPPRRFYLAGNQEQEFLQYQPQKQQGGTQSQKGKRQQ
EEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLGGENEEEEKGAIVTVKGGLSVISPTTEEQQQRPEEEEEK
PDCDEKDKHCQSQSRNGIDETICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSL
RKNAMFVPHYNLNANSIIYALNGRALVQVVCNNGERVFDGELQEGQVLIVPQNFAVAARSQSDNFEYVSF
KTNDRPSIGNLAGANSLLNALPEEVIQQTENLRRQARQVKNNNPFSFLVPPKESQRRVVA

Supporting Information S7 (Continued): Gly m 6 (Glycinin G3 Subunit) Consensus Sequence Generation

Gly m 6 G3 subunit Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
Gly m 6	Glycinin G3	(1)	MAKLVLSLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIESEGG	
Gly m 6 CAA33217.1 subunit G3		(1)	MAKLVLSLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIESEGG	
	Consensus	(1)	MAKLVLSLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIESEGG	
			51	100
Gly m 6	Glycinin G3	(51)	FIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIIYIQQSGIF	
Gly m 6 CAA33217.1 subunit G3		(51)	FIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIIYIQQSGIF	
	Consensus	(51)	FIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIIYIQQSGIF	
			101	150
Gly m 6	Glycinin G3	(101)	GMIFPGCPSTFEFPQQKGQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMY	
Gly m 6 CAA33217.1 subunit G3		(101)	GMIFPGCPSTFEFPQQKGQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMY	
	Consensus	(101)	GMIFPGCPSTFEFPQQKGQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMY	
			151	200
Gly m 6	Glycinin G3	(151)	NNEDTPVVAVSLIDTNSFQNQLDQMPRRFYLAGNQEQEFLQYQPQKQGG	
Gly m 6 CAA33217.1 subunit G3		(151)	NNEDTPVVAVSLIDTNSFQNQLDQMPRRFYLAGNQEQEFLQYQPQKQGG	
	Consensus	(151)	NNEDTPVVAVSLIDTNSFQNQLDQMPRRFYLAGNQEQEFLQYQPQKQGG	
			201	250
Gly m 6	Glycinin G3	(201)	TQSQKGKRQQUEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLQGENEEEE	
Gly m 6 CAA33217.1 subunit G3		(201)	TQSQKGKRQQUEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLQGENEEEE	
	Consensus	(201)	TQSQKGKRQQUEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLQGENEEEE	
			251	300
Gly m 6	Glycinin G3	(251)	KGAIVTVKGGLSVISPPTEEQQRPEEEEEKPDCEKDKHCQSQSRNGIDE	
Gly m 6 CAA33217.1 subunit G3		(251)	KGAIVTVKGGLSVISPPTEEQQRPEEEEEKPDCEKDKHCQSQSRNGIDE	
	Consensus	(251)	KGAIVTVKGGLSVISPPTEEQQRPEEEEEKPDCEKDKHCQSQSRNGIDE	
			301	350
Gly m 6	Glycinin G3	(301)	TICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSL	
Gly m 6 CAA33217.1 subunit G3		(301)	TICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSL	
	Consensus	(301)	TICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSL	
			351	400
Gly m 6	Glycinin G3	(351)	RKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIV	
Gly m 6 CAA33217.1 subunit G3		(351)	RKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIV	
	Consensus	(351)	RKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIV	
			401	450
Gly m 6	Glycinin G3	(401)	PQNFAVAARSQSDNFEYVSFKTNDRPSIGNLAGANSL LNALPEEVIQQT	
Gly m 6 CAA33217.1 subunit G3		(401)	PQNFAVAARSQSDNFEYVSFKTNDRPSIGNLAGANSL LNALPEEVIQQT	
	Consensus	(401)	PQNFAVAARSQSDNFEYVSFKTNDRPSIGNLAGANSL LNALPEEVIQQT	
			451	481
Gly m 6	Glycinin G3	(451)	NLRRQARQVKNNNPFSLVPPKESQRRVVA	
Gly m 6 CAA33217.1 subunit G3		(451)	NLRRQARQVKNNNPFSLVPPKESQRRVVA	
	Consensus	(451)	NLRRQARQVKNNNPFSLVPPKESQRRVVA	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 6 Glycinin G3 481aa

MAKLVLSLCFLLFSGCCFAFSFREQPQQNECQIQRLNALKPDNRIESEGGFIETWNPNNKPFQCAGVALSRCTLNRNALRRPSYTNAPQEIIYIQQSGIFGMIFPGCPSTFEFPQQKGQSSRPQDRHQKIYHFREGDLIAVPTGFAYWMYNNEDTPVVAVSLIDTNSFQNQLDQMPRRFYLAGNQEQEFLQYQPQKQGGTQSQKGKRQQUEENEGGSILSGFAPEFLEHAFVVDQRQIVRKLQGENEEEEKGAIVTVKGGLSVISPPTEEQQRPEEEEEKPDCEKDKHCQSQSRNGIDETICTMRLRHNIGQTSSPDIFNPQAGSITTATSLDFPALS WLKLSAQFGSLRKNAMFVPHYNLNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIVPQNFAVAARSQSDNFEYVSFKTNDRPSIGNLAGANSL LNALPEEVIQQTFLNLRQARQVKNNNPFSLVPPKESQRRVVA

Supporting Information S8: Gly m 6 (Glycinin G4 Subunit) Consensus Sequence Generation

Gly m 6 G4 subunit Protein Sequences:

>Gly m 6 Glycinin G4 562aa
MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVESEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHSPSYSPYPRMIIIAQKGKALGVAIPGCPETFEEPQEQSNRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPPSPVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLESPPDERKQIVTVEGGLSVISPKWQEQQ
DEDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDEKPRPSRPSQGKRNTGQDEDEDEDEDEQPRKS
REWRSKKTQPRRPRQEEPRERGCESTRNGVEENICTLKLHENIARPSRADFYNPAGRISTLNSLTLPALR
QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVAVFDGELRRGQLLVVPQNFVVAEQ
AGEQGFYIVFKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEKNWGPLVNPESQQGSPRVK
VA

>Gly m 6 CAA26478.1 562aa
MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVESEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHSPSYSPYPRMIIIAQKGKALGVAIPGCPETFEEPQEQSNRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPPSPVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLESPPDERKQIVTVEGGLSVISPKWQEQQ
DEDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDEKPRPSRPSQGKRNTGQDEDEDEDEDEQPRKS
REWRSKKTQPRRPRQEEPRERGCESTRNGVEENICTLKLHENIARPSRADFYNPAGRISTLNSLTLPALR
QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVAVFDGELRRGQLLVVPQNFVVAEQ
AGEQGFYIVFKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEKNWGPLVNPESQQGSPRVK
VA

>Gly m 6 CAA60533.1 563aa
MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVESEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHSPSYSPYPRMIIIAQKGKALGVAIPGCPETFEEPQEQSNRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPPGVVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLSPPDERKQIVTVEGGLSVISPKWQEQQ
DEDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDEKPRPSRPSHGKREQDQDQDEDEDEDEDEQPRK
SREWRSKKTQPRRPRQEEPRERGCESTRNGVEENICTLKLHENIARPSRADFYNPAGRISTLNSLTLPAL
RQFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVAVFDGELRRGQLLVVPQNFVVAE
QAGEQGFYIVFKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEKNWGPLVNPESQQGSPRV
KVA

>Gly m 6 UniProt Q9SB11 G4 subunit 563aa
MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVESEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHSPSYSPYPRMIIIAQKGKALGVAIPGCPETFEEPQEQSNRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPPGVVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLSPPDERKQIVTVEGGLSVISPKWQEQQ
DEDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDEKPRPSRPSQGKREQDQDQDEDEDEDEDEQPRK
SREWRSKKTQPRRPRQEEPRERGCESTRNGVEENICTLKLHENIARPSRADFYNPAGRISTLNSLTLPAL
RQFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVAVFDGELRRGQLLVVPQNFVVAE
QAGEQGFYIVFKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEKNWGPLVNPESQQGSPRV
KVA

>Gly m 6 CAA55977.1 517aa
MGKPFTLSLSSLCLLLLSSACFAITSSKFNECQLNNLNALEPDHRVESEGGLIETWNSQHPELQCAGVT
VSKRTLNRNGLHSPSYSPYPMIIVVQKGGAIGFAFPGCPETFEKPPQQSSRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPPGVVYWTYNTGDEPVVAISLLDTSNFNQLDQNPVIFYLAGNPDIEHPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLRSPDDERKQIVTVEGGLSVISPKWQEQED
EDEDEDEEYEQTPSYPPRRPSHGKHEDDEDEDEDEEDQPRPDHPPQRPSPRPEQQEPRGRGCQTRNGVEENI
CTMKLHENIARPSRADFYNPAGRISTLNSLTLPALRQFGLSAQYVVLYRNGIYSPHWNLNANSVIYVTR
GKGRVRVVNCQGNVAVFDGELRRGQLLVVPQNFVVAEQGGEQGLEYYVFKTHHNAVSSYIKDFRAIPSEV
LSNSYNLGQSQVRQLKYQGNSGPLVNP

>Gly m 6 AAA33964.1 516aa
MGKPFTLSLSSLCLLLLSSACFAITSSKFNECQLNNLNALEPDHRVESEGGLIETWNSQHPELQCAGVT
VSKRTLNRNGLHSPSYLPYPMIIVVQKGGAIGFAFPGCPETFEKPPQQSSRRGSRSSQKQQLQDSHQKIRH
FNEGDLVLIPLGVVYWTYNTGDEPVVAISPLDTSNFNQLDQNPVIFYLAGNPDIEHPETMQQQQQQKSH
GGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFTNTEDIAEKLRSPDDERKQIVTVEGGLSVISPKWQEQED
EDEDEDEEYGRTPSYPPRRPSHGKHEDDEDEDEDEEDQPRPDHPPQRPSPRPEQQEPRGRGCQTRNGVEENI
CTMKLHENIARPSRADFYNPAGRISTLNSLTLPALRQFGLSAQYVVLYRNGIYSPDWNLNANSVTMTRG
KGRVRVVNCQGNVAVFDGELRRGQLLVVPQNPVVAEQGGEQGLEYYVFKTHHNAVSSYIKDFRVIPSEVL
SNSYNLGQSQVRQLKYQGNSGPLVNP

Supporting Information S8: Gly m 6 (Glycinin G4 Subunit) Consensus Sequence Generation

Gly m 6 G4 subunit Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(1)	MGKP-FTLSLSSSLCLLLLSSACFAISSKLNECQLNNLNALEPDHRSVESE
	Gly m 6	CAA26478.1	(1)	MGKP-FTLSLSSSLCLLLLSSACFAISSKLNECQLNNLNALEPDHRSVESE
	Gly m 6	CAA60533.1	(1)	MGKP-FTLSLSSSLCLLLLSSACFAISSKLNECQLNNLNALEPDHRSVESE
	Gly m 6	CAA55977.1	(1)	MGKPF-FTLSLSSSLCLLLLSSACFAISSKFNECQLNNLNALEPDHRSVESE
	Gly m 6	AAA33964.1	(1)	MGKPF-FTLSLSSSLCLLLLSSACFAISSKFNECQLNNLNALEPDHRSVESE
	Consensus		(1)	MGKPF-FTLSLSSSLCLLLLSSACFAISSKLNECQLNNLNALEPDHRSVESE
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(50)	GGLIQTWNSQHPELKCAGVTVSKLTNLRNGLHSPSYSPYPRMIITIAQGKG
	Gly m 6	CAA26478.1	(50)	GGLIQTWNSQHPELKCAGVTVSKLTNLRNGLHSPSYSPYPRMIITIAQGKG
	Gly m 6	CAA60533.1	(50)	GGLIQTWNSQHPELKCAGVTVSKLTNLRNGLHSPSYSPYPRMIITIAQGKG
	Gly m 6	CAA55977.1	(51)	GGLIETWNSQHPELKCAGVTVSKRTNLRNGLHSPSYSPYPMIITVQGKG
	Gly m 6	AAA33964.1	(51)	GGLIETWNSQHPELKCAGVTVSKRTNLRNGLHSPSYLPYPMIITVQGKG
	Consensus		(51)	GGLIQTWNSQHPELKCAGVTVSKLTNLRNGLHSPSYSPYPRMIITIAQGKG
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(100)	ALGVAIPGCPETFEPEQEQSNRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
	Gly m 6	CAA26478.1	(100)	ALGVAIPGCPETFEPEQEQSNRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
	Gly m 6	CAA60533.1	(100)	ALGVAIPGCPETFEPEQEQSNRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
	Gly m 6	CAA55977.1	(101)	ALGFAPGCPETFEKPKQQSSRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
	Gly m 6	AAA33964.1	(101)	ALGFAPGCPETFEKPKQQSSRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
	Consensus		(101)	ALGVAIPGCPETFEPEQEQSNRRGSRSQKQQLQDSSHQKIRHFNEDVLVI
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(150)	PSPVPYWTYNTGDEPVVAISLLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
	Gly m 6	CAA26478.1	(150)	PSPVPYWTYNTGDEPVVAISLLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
	Gly m 6	CAA60533.1	(150)	PSPVPYWTYNTGDEPVVAISLLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
	Gly m 6	CAA55977.1	(150)	PSPVPYWTYNTGDEPVVAISLLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
	Gly m 6	AAA33964.1	(150)	PLGVVPYWTYNTGDEPVVAISPLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
	Consensus		(151)	PSPVPYWTYNTGDEPVVAISLLDTSNFFNNQLDQTPRVFYLAGNPDIETPE
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(200)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
	Gly m 6	CAA26478.1	(200)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
	Gly m 6	CAA60533.1	(200)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
	Gly m 6	CAA55977.1	(200)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
	Gly m 6	AAA33964.1	(200)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
	Consensus		(201)	TMQQQQQQKSHGGRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIA
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(250)	EKLESPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSH
	Gly m 6	CAA26478.1	(250)	EKLESPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSH
	Gly m 6	CAA60533.1	(250)	EKLQSPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSH
	Gly m 6	CAA55977.1	(249)	EKLQSPDDERKQIVTVEGGLSVISPKWQEQ--EDEDDEDEDEEYEQTPSY
	Gly m 6	AAA33964.1	(249)	EKLQSPDDERKQIVTVEGGLSVISPKWQEQ--EDEDDEDEDEEYGRTPSY
	Consensus		(251)	EKLQSPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSH
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(300)	PRRPSHGKREQDEDEDEDEDEDKPRPSRPSQGRN-KTGQDEDEDEDEDEQPR
	Gly m 6	CAA26478.1	(300)	PRRPSHGKREQDEDEDEDEDEDKPRPSRPSQGRN-KTGQDEDEDEDEDEQPR
	Gly m 6	CAA60533.1	(300)	PRRPSHGKREQDEDEDEDEDEDKPRPSRPSHGKREQDQDQDEDEDEDEDEQPR
	Gly m 6	CAA55977.1	(297)	PRRPSHGKHEDDEDEDEDEDEQPRPDH-----PQ
	Gly m 6	AAA33964.1	(297)	PRRPSHGKHEDDEDEDEDEDEQPRPDH-----PQ
	Consensus		(301)	PRRPSHGKREQDEDEDEDEDEDKPRPSRPSQGRN-KTGQDEDEDEDEDEQPR
Gly m 6 UniProt Q9SB11	Gly m 6	Glycinin G4	(349)	KSREWRSKKTQPRRPQEPERERGCTRNNGVEENICTLKLHENIARPSRA
	Gly m 6	CAA26478.1	(349)	KSREWRSKKTQPRRPQEPERERGCTRNNGVEENICTLKLHENIARPSRA
	Gly m 6	CAA60533.1	(350)	KSREWRSKKTQPRRPQEPERERGCTRNNGVEENICTLKLHENIARPSRA
	Gly m 6	CAA55977.1	(326)	R-----PSRPEQEPGRGRCQTRNNGVEENICTMKLHENIARPSRA
	Gly m 6	AAA33964.1	(326)	R-----PSRPEQEPGRGRCQTRNNGVEENICTMKLHENIARPSRA
	Consensus		(351)	KSREWRSKKTQPRRPQEPERERGCTRNNGVEENICTLKLHENIARPSRA

Supporting Information S8: Gly m 6 (Glycinin G4 Subunit) Consensus Sequence Generation

			401		450
Gly m 6	Glycinin G4	(399)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
Gly m 6	CAA26478.1	(399)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
Gly m 6	CAA60533.1	(400)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
Gly m 6 UniProt Q9SB11	G4 subunit	(400)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
Gly m 6	CAA55977.1	(366)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
Gly m 6	AAA33964.1	(366)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
	Consensus	(401)	DFYNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSV		
			451		500
Gly m 6	Glycinin G4	(449)	IYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFEY		
Gly m 6	CAA26478.1	(449)	IYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFEY		
Gly m 6	CAA60533.1	(450)	IYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFEY		
Gly m 6 UniProt Q9SB11	G4 subunit	(450)	IYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFEY		
Gly m 6	CAA55977.1	(416)	IYVTRGKGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGLE		
Gly m 6	AAA33964.1	(416)	TMTR-GKGRVRVVCQGNVFDGELRRGQLLVVPQNPVVAEQAGEQGLE		
	Consensus	(451)	IYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFEY		
			501		550
Gly m 6	Glycinin G4	(499)	IVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLV		
Gly m 6	CAA26478.1	(499)	IVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLV		
Gly m 6	CAA60533.1	(500)	IVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLV		
Gly m 6 UniProt Q9SB11	G4 subunit	(500)	IVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLV		
Gly m 6	CAA55977.1	(466)	VVFKTHHNAVSSYIKDVFRVIPSEVLSNSYNLGSQVRLKYQGNSSGPLV		
Gly m 6	AAA33964.1	(465)	VVFKTHHNAVSSYIKDVFRVIPSEVLSNSYNLGSQVRLKYQGNSSGPLV		
	Consensus	(501)	IVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLV		
			551		564
Gly m 6	Glycinin G4	(549)	NPESQQGSPRVKVA		
Gly m 6	CAA26478.1	(549)	NPESQQGSPRVKVA		
Gly m 6	CAA60533.1	(550)	NPESQQGSPRVKVA		
Gly m 6 UniProt Q9SB11	G4 subunit	(550)	NPESQQGSPRVKVA		
Gly m 6	CAA55977.1	(516)	NP-----		
Gly m 6	AAA33964.1	(515)	NP-----		
	Consensus	(551)	NPESQQGSPRVKVA		

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 6 Glycinin G4 562aa
 MGKPFTLSLSSLCLLLSSACFAISSSKLNECQLNNLNALEPDHRVESEGGLIQTWNSQHPELKCAGVTV
 SKLTLNRNGLHSPSYSPYPRMIIIAQKGKALGVAIPGCPETFEEPQEQSNRRGSRSQKQLQDSHOKIRH
 FNEGDLVLPSPVPYWTYNTGDEPVVAISLLDTSNFNQQLDQTPRVFYLAGNPDIETMQQQQQKSH
 GGRKQGHQEEEEEGSVLSGFSKHFLAQSFNTNEDIAEKLESPDDERKQIVTVEGGLSVISPKWQEQQ
 DEDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDKPRPSRPSQGRNKTGQDEDEDEDEDQPRKS
 REWRSKKTQPRRPRQEEPRERGCETRNGVEENICTLKLHENIARPSRADFYNPKAGRISTLNSLTLPALR
 QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQ
 AGEQGFEYIVFKTHHNAVTSYIKDVFRVIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNPESQQGSPRVK
 VA

Supporting Information S9: Gly m 6 (Glycinin Precursor Subunit) Consensus Sequence Generation

Gly m 6 precursor Protein Sequences:

>Gly m 6 CAA37044.1 562aa
MGKPFTLSLSSLCLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHLPSPYPRMIIIAQKGKALQCKPGCPETFEEPQEQSNRRGSRQKQQLQDSHQKIRHF
NEGDLVIPGPVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQKSHG
GRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEKLQSPDDERKQIVTVEGGLSVISPKWQEQD
EDEDEDEDDEDEQIPSHPPRRPSHGKREQDEDEDEDKPRPSRPSQGKREQDQDQDEDEDEDQPRKS
REWRSKKTQPRRPRQEEPRERCETRNGVEENICTLKLHENIARPSRADFYNNPKAGRISTLNSLTLPALR
QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVFDGELRRGQLLVVPQNFVVAEQ
AGEQGFYIVFKTHHNAVTSYLDKDVFRAPSEVLAHSYNLRQSQVSELKYEGNWGPLVNPESQQGSPRVK
VA

>Gly m 6 Glycinin precursor 562aa
MGKPFTLSLSSLCLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHLPSPYPRMIIIAQKGKALQCKPGCPETFEEPQEQSNRRGSRQKQQLQDSHQKIRHF
NEGDLVIPGPVYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIEYPETMQQQQQKSHG
GRKQGQHQQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEKLQSPDDERKQIVTVEGGLSVISPKWQEQD
EDEDEDEDDEDEQIPSHPPRRPSHGKREQDEDEDEDKPRPSRPSQGKREQDQDQDEDEDEDQPRKS
REWRSKKTQPRRPRQEEPRERCETRNGVEENICTLKLHENIARPSRADFYNNPKAGRISTLNSLTLPALR
QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVNCQGNVFDGELRRGQLLVVPQNFVVAEQ
AGEQGFYIVFKTHHNAVTSYLDKDVFRAPSEVLAHSYNLRQSQVSELKYEGNWGPLVNPESQQGSPRVK
VA

Supporting Information S9 (Continued): Gly m 6 (Glycinin Precursor Subunit) Consensus Sequence Generation

Gly m 6 precursor Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
Gly m 6	Gly m 6 CAA37044.1	(1)	MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEG	
Gly m 6	Glycinin precursor	(1)	MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEG	
	Consensus	(1)	MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEG	
			51	100
Gly m 6	Gly m 6 CAA37044.1	(51)	GLIQTWNSQHPELKCAGVTVSKLTLNRNGLHLPSPYPRMIIIAQGKGA	
Gly m 6	Glycinin precursor	(51)	GLIQTWNSQHPELKCAGVTVSKLTLNRNGLHLPSPYPRMIIIAQGKGA	
	Consensus	(51)	GLIQTWNSQHPELKCAGVTVSKLTLNRNGLHLPSPYPRMIIIAQGKGA	
			101	150
Gly m 6	Gly m 6 CAA37044.1	(101)	LQCKPGCPETFEEPQEESNRRGSRSSQKQLQDSHQKIRHFNEDVLIIPP	
Gly m 6	Glycinin precursor	(101)	LQCKPGCPETFEEPQEESNRRGSRSSQKQLQDSHQKIRHFNEDVLIIPP	
	Consensus	(101)	LQCKPGCPETFEEPQEESNRRGSRSSQKQLQDSHQKIRHFNEDVLIIPP	
			151	200
Gly m 6	Gly m 6 CAA37044.1	(151)	GVPHYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIIEYPETM	
Gly m 6	Glycinin precursor	(151)	GVPHYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIIEYPETM	
	Consensus	(151)	GVPHYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIIEYPETM	
			201	250
Gly m 6	Gly m 6 CAA37044.1	(201)	QQQQQQKSHGGRKQGQHQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEK	
Gly m 6	Glycinin precursor	(201)	QQQQQQKSHGGRKQGQHQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEK	
	Consensus	(201)	QQQQQQKSHGGRKQGQHQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEK	
			251	300
Gly m 6	Gly m 6 CAA37044.1	(251)	LQSPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSHPPR	
Gly m 6	Glycinin precursor	(251)	LQSPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSHPPR	
	Consensus	(251)	LQSPDDERKQIVTVEGGLSVISPKWQEQQDEDEDEDEDEDEDEQIPSHPPR	
			301	350
Gly m 6	Gly m 6 CAA37044.1	(301)	RPSHGKREQDEDEDEDEDEDEKPRPSRPSQKREQDQDQDEDEDEDEDEQPRKS	
Gly m 6	Glycinin precursor	(301)	RPSHGKREQDEDEDEDEDEDEKPRPSRPSQKREQDQDQDEDEDEDEDEQPRKS	
	Consensus	(301)	RPSHGKREQDEDEDEDEDEDEKPRPSRPSQKREQDQDQDEDEDEDEDEQPRKS	
			351	400
Gly m 6	Gly m 6 CAA37044.1	(351)	REWRSKKTQPRRPRQEEPRERGCETRNGVEENICTLKLHENIARPSRADF	
Gly m 6	Glycinin precursor	(351)	REWRSKKTQPRRPRQEEPRERGCETRNGVEENICTLKLHENIARPSRADF	
	Consensus	(351)	REWRSKKTQPRRPRQEEPRERGCETRNGVEENICTLKLHENIARPSRADF	
			401	450
Gly m 6	Gly m 6 CAA37044.1	(401)	YNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSVIY	
Gly m 6	Glycinin precursor	(401)	YNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSVIY	
	Consensus	(401)	YNPKAGRISTLNSLTLPALRQFQLSAQYVVLYKNGIYSPHWNLNANSVIY	
			451	500
Gly m 6	Gly m 6 CAA37044.1	(451)	VTRGQGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFYIV	
Gly m 6	Glycinin precursor	(451)	VTRGQGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFYIV	
	Consensus	(451)	VTRGQGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQAGEQGFYIV	
			501	550
Gly m 6	Gly m 6 CAA37044.1	(501)	FKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNP	
Gly m 6	Glycinin precursor	(501)	FKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNP	
	Consensus	(501)	FKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNP	
			551	562
Gly m 6	Gly m 6 CAA37044.1	(551)	ESQQGSPRVKVA	
Gly m 6	Glycinin precursor	(551)	ESQQGSPRVKVA	
	Consensus	(551)	ESQQGSPRVKVA	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>Gly m 6 Glycinin precursor 562aa

MGKPFTLSLSSLCLLLLSSACFAISSSKLNECQLNNLNALEPDHRVEFEGGLIQTWNSQHPELKCAGVTV
SKLTLNRNGLHLPSPYPRMIIIAQKGALQCKPGCPETFEEPQEESNRRGSRSSQKQLQDSHQKIRHF
NEGDVLIIPGVPHYWTYNTGDEPVVAISLLDTSNFNQLDQTPRVFYLAGNPDIIEYPETMQQQQQKSHG
GRKQGQHQEEEEEGGSVLSGFSKHFLAQSFNTNEDIAEKLQSPDDERKQIVTVEGGLSVISPKWQEQQD
EDEDEDEDEDEQIPSHPPRRPSHGKREQDEDEDEDEDEKPRPSRPSQKREQDQDQDEDEDEDEDEQPRKS
REWRSKKTQPRRPRQEEPRERGCETRNGVEENICTLKLHENIARPSRADFYNPAGRISTLNSLTLPALR
QFQLSAQYVVLYKNGIYSPHWNLNANSVIYVTRGQGKVRVVCQGNVFDGELRRGQLLVVPQNFVVAEQ
AGEQGFYIVFKTHHNAVTSYLDKDFRAIPSEVLAHSYNLRQSQVSELKYEGNWGPLVNPESQQGSPRVK
VA

Supporting Information S10: Gly m Bd 28 K Consensus Sequence Generation

Gly m Bd 28K Protein Sequences:

>28K BAB21619 473aa
KTTLLLLLVFLCHGVATTTMAFHDDEGGDKKSPKSLFLMSNSTRVFKTDAGEMRVLKSHGGRIFYRHMHI
GFISMPEKSLFVPQYLDNSLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLH
VICSIDPSTSLGLETQSFYIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDDSHA
PSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYNLYD
DKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGYPNG
SRAMKTKIKQGDVFVVPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASLLRTLTMGPELSAAFGV
SEDTLRRAVDAQHEAVILPSAWAAPPENAGKLKMEEEPNAIRSFANDVVMDF

>28K 12697782 476aa
MGNKTTTTLLLVFLCHGVATTTMAFHDDEGGDKKSPKSLFLMSNSTRVFKTDAGEMRVLKSHGGRIFYRH
MHIGFISMPEKSLFVPQYLDNSLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQ
RLHVICSIDPSTSLGLETQSFYIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDD
SHAPSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYN
LYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY
PNGSRAMKTKIKQGDVFVVPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASLLRTLTMGPELSAA
FGVSEDTLRRAVDAQHEAVILPSAWAAPPENAGKLKMEEEPNAIRSFANDVVMDF

>28K ACD36978.1 455aa
VLCHGVATTTMAFHDDEGGDKKSPKSLFLMSNSTRVFKTDAGEMRVLKSHGGRIFYRHMHIGFISMPEKS
LFVPQYLDNSLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDPST
SLGLETQSFYIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDDSHAPSLWTKFLQ
LKKDDKEQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNA
YGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGYPNGSKAMKTKIK
QGDVFVVPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASLLRTLTMGPELSAAFGVSEDTLRRAV
DAQHEAVILPSAWAAPRKMQEAEMEESQMLLKLCQ

>28K ACD36975.1 373aa
LDSNLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDPSTSLGLET
FQSFYIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDDSHAPSLWTKFLQLKKDDK
EQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKA
LHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFV
VPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASLLRTLTMGPELSAAFGVSEDTLRRAVDAQHEA
VILPSAWAAPPENAGKLKMEEEP

>28K ACD36976.1 373aa
LDSNLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDPSTSLGLET
FQSFNIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDDSHAPSLWTKFLQLKKDDK
EQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKA
LHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFV
VPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASLLRTLTMGPELSAAFGVSEDTLRRAVDAQHAA
VILPSAWAAPPENAGKLKMEEEP

>28K ACD36974.1 320aa
LDSNLIIFIRRGAEAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDPSTSLGLET
FQSFYIGGGANSHSVLSGFEPAALETAFNESRTVVEEIFSKELDGPIMFVDDSHVPSLWTKFLQLKKDDK
EQQLKKMMQDQEEDEEEKQTSRSWRKLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKA
LHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFV
VPYRYPFCQVASRDGPLEFFGFSTSAKKNKPQFLAGAASL

Supporting Information S10 (Continued): Gly m Bd 28 K Consensus Sequence Generation

Gly m Bd 28K Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1		50
28K	BAB21619	(1)	---KTLLLLLLF	VLCHGVATTMTMAFHDDDEGGDKKSPKSLFLMSNSTRVFK	
28K	12697782	(1)	MGNKTTLLLLLLF	VLCHGVATTMTMAFHDDDEGGDKKSPKSLFLMSNSTRVFK	
28K	ACD36978.1	(1)	-----	VLCHGVATTMTMAFHDDDEGGDKKSPKSLFLMSNSTRVFK	
28K	ACD36975.1	(1)	-----		
28K	ACD36976.1	(1)	-----		
28K	ACD36974.1	(1)	-----		
	Consensus	(1)		VLCHGVATTMTMAFHDDDEGGDKKSPKSLFLMSNSTRVFK	
			51		100
28K	BAB21619	(48)	TDAGEMRVLKSHGGRIFYRHHIGFISMEPKSLFVPQY	LDSNLIIFIRRG	
28K	12697782	(51)	TDAGEMRVLKSHGGRIFYRHHIGFISMEPKSLFVPQY	LDSNLIIFIRRG	
28K	ACD36978.1	(39)	TDAGEMRVLKSHGGRIFYRHHIGFISMEPKSLFVPQY	LDSNLIIFIRRG	
28K	ACD36975.1	(1)	-----	LDSNLIIFIRRG	
28K	ACD36976.1	(1)	-----	LDSNLIIFIRRG	
28K	ACD36974.1	(1)	-----	LDSNLIIFIRRG	
	Consensus	(51)	TDAGEMRVLKSHGGRIFYRHHIGFISMEPKSLFVPQY	LDSNLIIFIRRG	
			101		150
28K	BAB21619	(98)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
28K	12697782	(101)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
28K	ACD36978.1	(89)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
28K	ACD36975.1	(13)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
28K	ACD36976.1	(13)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
28K	ACD36974.1	(13)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
	Consensus	(101)	EAKLGFIYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQRLHVICSIDP		
			151		200
28K	BAB21619	(148)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
28K	12697782	(151)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
28K	ACD36978.1	(139)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
28K	ACD36975.1	(63)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
28K	ACD36976.1	(63)	STSLGLETQSFN	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	TFSKE
28K	ACD36974.1	(63)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
	Consensus	(151)	STSLGLETQSFY	IGGGANSHSVLSGFEPAILTAFNESRTVVEE	IFSKE
			201		250
28K	BAB21619	(198)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
28K	12697782	(201)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
28K	ACD36978.1	(189)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
28K	ACD36975.1	(113)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
28K	ACD36976.1	(113)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
28K	ACD36974.1	(113)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
	Consensus	(201)	LDGPIMFVDDSH	APSLWTKFLQLKKDDKEQQLKKMMQDQEEDEEEKQTSR	
			251		300
28K	BAB21619	(248)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
28K	12697782	(251)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
28K	ACD36978.1	(239)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
28K	ACD36975.1	(163)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
28K	ACD36976.1	(163)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
28K	ACD36974.1	(163)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
	Consensus	(251)	SWRKLLLETVFGKVNEKIENKDTAGSPASYNLYDDKKADFKNAYGWSKALH		
			301		350
28K	BAB21619	(298)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
28K	12697782	(301)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
28K	ACD36978.1	(289)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
28K	ACD36975.1	(213)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
28K	ACD36976.1	(213)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
28K	ACD36974.1	(213)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		
	Consensus	(301)	GGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY		

Supporting Information S10 (Continued): Gly m Bd 28 K Consensus Sequence Generation

			351	400
28K	BAB21619	(348)	PNGSRAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
28K	12697782	(351)	PNGSRAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
28K	ACD36978.1	(339)	PNGSKAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
28K	ACD36975.1	(263)	PNGSKAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
28K	ACD36976.1	(263)	PNGSKAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
28K	ACD36974.1	(263)	PNGSKAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
	Consensus	(351)	PNGSKAMKTKIKQGDVFVVP	RYFPFCQVASRDGPLEFFGFST
			401	450
28K	BAB21619	(398)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
28K	12697782	(401)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
28K	ACD36978.1	(389)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
28K	ACD36975.1	(313)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
28K	ACD36976.1	(313)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
28K	ACD36974.1	(313)	FLAGAASLL	-----
	Consensus	(401)	FLAGAASLLRTL	MGPELSAAFGVSEDTLRR
			451	476
28K	BAB21619	(448)	NAGKLKMEEEPNA	IRSFANDVVM
28K	12697782	(451)	NAGKLKMEEEPNA	IRSFANDVVM
28K	ACD36978.1	(438)	KMQEAE	MEESQMLLKL
28K	ACD36975.1	(363)	NAGKLKMEEEP	-----
28K	ACD36976.1	(363)	NAGKLKMEEEP	-----
28K	ACD36974.1	(321)	-----	-----
	Consensus	(451)	NAGKLKMEEEP	IR

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>28K 12697782 476aa
 MGNTTLLLLLVLCHGVATTMAFHDDDEGGDKKSPKSLFLMSNSTRVFKTDAGEMRVLKSHGGRIFYRH
 MHIGFISMEPKSLFVPQYLDNLIFIRRGAKLGFYDDELAERRLKTGDLYMIPSGSAFYLVNIGEGQ
 RLHVICSIDPSTSLGLETFQSFYIGGGANSHSVLSGFEPAILETAFTNESRTVVVEIFSKELDGPIMFVDD
 SHAPSLWTKFLQLKKDDKEQQLKMMQDQEEDEEEKQTSRSWRKLLTVFGKVNEKIENKDTAGSPASYN
 LYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLAPHVNPISDEYTIVLSGYGELHIGY
 PNGSRAMKTKIKQGDVFVVP

Supporting Information S11: Gly m Bd 30 K Consensus Sequence Generation

Gly m Bd 30K Protein Sequences:

>30K ABC56139 379aa
MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NLNYIRDMNANRKSPHSHRLGLNKFADITPQEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGSGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGcYNGWHYQSFewVL
EHGGIATDDDDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLVsARVKGHRRVDHSPL

>30K 84371705 379aa
MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NLNYIRDMNANRKSPHSHRLGLNKFADITPQEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGSGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGcYNGWHYQSFewVL
EHGGIATDDDDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLVsARVKGHRRVDHSPL

>30K BAA25899.1 379aa
MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NLNYIRDMNANRKSPHSHRLGLNKFADITPQEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGSGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGcYNGWHYQSFewVL
EHGGIATDDDDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLVsARVKGHRRVDHSPL

>30K AAB09252.1 379aa
MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NSNYIRDMNANRKSPHSHRLGLNKFADITPQEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGRGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGsYNGWQYQSFewVL
EHGGIATDDDDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGEDWGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLVsARVKGHRRVDHSPL

>30K P22895.1 379aa
MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NSNYIRDMNANRKSPHSHRLGLNKFADITPQEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGRGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGsYNGWQYQSFewVL
EHGGIATDDDDYPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSADGVDYWIAKNSWGFWDGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLVsARVKGHRRVDHSPL

Gly m Bd 30K Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

			1	50
30K	ABC56139	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
30K	84371705	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
30K	BAA25899.1	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
30K	AAB09252.1	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
30K	P22895.1	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
	Consensus	(1)	MGFLVLLLFSLGLSSSSSISTHRSILDLDLTkFTTQKQVSSLFQLWKSE	
			51	100
30K	ABC56139	(51)	HGRVYHNHEEEAKRLEIFKNNLNYIRDMNANRKSPHSHRLGLNKFADITP	
30K	84371705	(51)	HGRVYHNHEEEAKRLEIFKNNLNYIRDMNANRKSPHSHRLGLNKFADITP	
30K	BAA25899.1	(51)	HGRVYHNHEEEAKRLEIFKNNLNYIRDMNANRKSPHSHRLGLNKFADITP	
30K	AAB09252.1	(51)	HGRVYHNHEEEAKRLEIFKNNsNYIRDMNANRKSPHSHRLGLNKFADITP	
30K	P22895.1	(51)	HGRVYHNHEEEAKRLEIFKNNsNYIRDMNANRKSPHSHRLGLNKFADITP	
	Consensus	(51)	HGRVYHNHEEEAKRLEIFKNNLNYIRDMNANRKSPHSHRLGLNKFADITP	
			101	150
30K	ABC56139	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
30K	84371705	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
30K	BAA25899.1	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
30K	AAB09252.1	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
30K	P22895.1	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
	Consensus	(101)	QEFsSKKYLQAPKDVSSQIKMANKKMKKEQYSCDHPPASWDWRKKGVITQV	
			151	200
30K	ABC56139	(151)	KYQGGCGSGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGcYNG	
30K	84371705	(151)	KYQGGCGSGWAFSATGAIEAAHAIAITGDLVSLSEQELVDCVEESEGcYNG	

Supporting Information S11 (Continued): Gly m Bd 30 K Consensus Sequence Generation

30K	BAA25899.1	(151)	KYQGGCGSGWAFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEG	CYNG
30K	AAB09252.1	(151)	KYQGGCGRGWAFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEG	SYNG
30K	P22895.1	(151)	KYQGGCGRGWAFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEG	SYNG
	Consensus	(151)	KYQGGCGSGWAFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEG	CYNG
			201	250
30K	ABC56139	(201)	WHYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
30K	84371705	(201)	WHYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
30K	BAA25899.1	(201)	WHYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
30K	AAB09252.1	(201)	WQYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
30K	P22895.1	(201)	WQYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
	Consensus	(201)	WHYQSFEWVLEHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMS	
			251	300
30K	ABC56139	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
30K	84371705	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
30K	BAA25899.1	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
30K	AAB09252.1	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
30K	P22895.1	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
	Consensus	(251)	DESTESETEQAFLSAILEQPISVSIDAKDFHLYTGGIYDGENCTSPYGIN	
			301	350
30K	ABC56139	(301)	HFVLLVGYGSAADGVYWIAKNSWG	EDWGEDGYIWIQRNTGNLLGVCGMNY
30K	84371705	(301)	HFVLLVGYGSAADGVYWIAKNSWG	EDWGEDGYIWIQRNTGNLLGVCGMNY
30K	BAA25899.1	(301)	HFVLLVGYGSAADGVYWIAKNSWG	EDWGEDGYIWIQRNTGNLLGVCGMNY
30K	AAB09252.1	(301)	HFVLLVGYGSAADGVYWIAKNSWG	EDWGEDGYIWIQRNTGNLLGVCGMNY
30K	P22895.1	(301)	HFVLLVGYGSAADGVYWIAKNSWG	FEDWGEDGYIWIQRNTGNLLGVCGMNY
	Consensus	(301)	HFVLLVGYGSAADGVYWIAKNSWG	EDWGEDGYIWIQRNTGNLLGVCGMNY
			351	379
30K	ABC56139	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	
30K	84371705	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	
30K	BAA25899.1	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	
30K	AAB09252.1	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	
30K	P22895.1	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	
	Consensus	(351)	FASYPTKEESETLV SARVKGHRRVDHSPL	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>30K 84371705 379aa
MGFLVLLFSLGLSSSSSISTHRSILDLDLTKFTTQKQVSSLFQLWKSEHGRVYHNHEEEAKRLEIFKN
NLNYIRDMNANRKS PHSRLGLNKFADITPQEFSSKKYLQAPKDVSQIKMANKMKKEQYSCDHPPASWD
WRKKGVITQVKYQGGCGSGWAFSATGAIEAAHAIATGDLVSLSEQELVDCVEESEG CYNGWHYQSFEWVL
EHGGIATDDDPYRAKEGRCKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPISVSIDAKDF
HLYTGGIYDGENCTSPYGINHFVLLVGYGSAADGVYWIAKNSWGEDWGEDGYIWIQRNTGNLLGVCGMNY
FASYPTKEESETLV SARVKGHRRVDHSPL

Supporting Information S12: Kunitz Trypsin Inhibitor 1 Consensus Sequence Generation

Kunitz Trypsin Inhibitor 1 Protein Sequences:

>KTI 1 203aa
MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRGKGGGIEVDSTGKEICPLTVV
QSPNELDKIGLVFTSPLHALFIAERYPLSIKFGSFAVITLCAGMPTEWAIVEREGLQAVKLAARDTVDG
WFNIERSVREYNDYKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLSKNKPLVVQFQKFRSSTA

>KTI 1 AAB23482.1 203aa
MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRGKGGGIEVDSTGKEICPLTVV
QSPNELDKIGLVFTSPLHALFIAERYPLSIKFGSFAVITLCAGMPTEWAIVEREGLQAVKLAARDTVDG
WFNIERSVREYNDYKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLSKNKPLVVQFQKFRSSTA

>KTI 1 AAB23483.1 204aa
MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRGKSGGIEGNSTGKEICPLTVV
QSPNKHNGIGLVFKSPLHALFIAERYPLSIKFGSFAVIPLCGVMPTKWAIVEREGLQAVTLAARDTVDG
WFNIERSVREYNDYKLVFCPQQAEDNKCEDIGIQIDNDGIRRLVLSKNKPLVVEFQKFRSSTA

>KTI 1 CAA56343.1 208aa
MKSTTSLALFLLCAL TSSYQPSATADIVFDTEGNPIRNGGTY YLPVIRGKGGGIEFAKTETETCPLTVV
QSPFEGLRGLPLIISSPFKILDITEGLILSLK FHLCTPLSLNSFSVD RYSQGSARRTPCQTHWLQKHNR
CWFRIQRASSENYYKLVFCTSNDDSSCGDIVAPIDREGNRPLIVTHDQNHPLL VQFQKVEAYESSTA

Kunitz Trypsin Inhibitor 1 Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

		1		50
KTI 1	(1)	MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRG		
KTI 1 AAB23482.1	(1)	MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRG		
KTI 1 AAB23483.1	(1)	MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRG		
KTI 1 CAA56343.1	(1)	MKSTTSLALFLLCAL TSSYQPSATADIVFDTEGNPIRNGGTY YLPVIRG		
Consensus	(1)	MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRG		
		51		100
KTI 1	(51)	KGGGIEVDSTGKEICPLTVVQSPNE-LDKGIGLVFTSPLHALFIAERYPL		
KTI 1 AAB23482.1	(51)	KGGGIEVDSTGKEICPLTVVQSPNE-LDKGIGLVFTSPLHALFIAERYPL		
KTI 1 AAB23483.1	(51)	KSGGIEGNSTGKEICPLTVVQSPNK-HNKGIGLVFKSPLHALFIAERYPL		
KTI 1 CAA56343.1	(51)	KGGGIEFAKTETETCPLTVVQSPFEGLRGLPLIISSPFKILDITEGLIL		
Consensus	(51)	KGGGIEVDSTGKEICPLTVVQSPNE LDKGIGLVFTSPLHALFIAERYPL		
		101		150
KTI 1	(100)	SIKFGSFAVITLCAGMPTEWAIVEREGLQAVKLAARDTVDGWFNIERSR		
KTI 1 AAB23482.1	(100)	SIKFGSFAVITLCAGMPTEWAIVEREGLQAVKLAARDTVDGWFNIERSR		
KTI 1 AAB23483.1	(100)	SIKFD SFAVIPLCGVMPTKWAIVEREGLQAVTLAARDTVDGWFNIERSR		
KTI 1 CAA56343.1	(101)	SLK FHLCTPLSLNSFSVD RYSQGSARRTPCQTHWLQKHNR CWFRIQRASS		
Consensus	(101)	SIKFGSFAVITLCAGMPTEWAIVEREGLQAVTLAARDTVDGWFNIERSR		
		151		200
KTI 1	(150)	EYND-YKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLS--KNKPLVVQFQ		
KTI 1 AAB23482.1	(150)	EYND-YKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLS--KNKPLVVQFQ		
KTI 1 AAB23483.1	(150)	EYNDYKLVFCPQQAEDNKCEDIGIQIDNDGIRRLVLS--KNKPLVVEFQ		
KTI 1 CAA56343.1	(151)	ESNY-YKLVFCTSN-DDSSCGDIVAPIDREGNRPLIVTHDQNHPLL VQFQ		
Consensus	(151)	EYND YKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLS KNKPLVVQFQ		
		201		
KTI 1	(197)	KFRSSTA---		
KTI 1 AAB23482.1	(197)	KFRSSTA---		
KTI 1 AAB23483.1	(198)	KFRSSTA---		
KTI 1 CAA56343.1	(199)	KVEAYESSTA		
Consensus	(201)	KFRSSTA		

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>KTI 1 203aa
MKSTIFFALFLVCAFTISYLP SATAQFVLDTDDDP LQNGGTY YMLPVMRGKGGGIEVDSTGKEICPLTVV
QSPNELDKIGLVFTSPLHALFIAERYPLSIKFGSFAVITLCAGMPTEWAIVEREGLQAVKLAARDTVDG
WFNIERSVREYNDYKLVFCPQQAEDNKCEDIGIQIDDDGIRRLVLSKNKPLVVQFQKFRSSTA

Supporting Information S13: Kunitz Trypsin Inhibitor 3 Consensus Sequence Generation

>KTI 3 216aa

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITAFGGIRAAPTGNERCPLTVVQS
RNELDKGIGTISSPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDG
WFLRLERVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESLAKKNH
GLSRSE

>KTI 3 AAB23464.1 216aa

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITAFGGIRAAPTGNERCPLTVVQS
RNELDKGIGTISSPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDG
WFLRLERVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESLAKKNH
GLSRSE

>KTI 3 CAA45777.1 217aa

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITAFGGIRAAPTGNERCPLTVVQ
SRNELDKGIGTISSPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMD
GWFLRLERVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESLAKKN
HGLSRSE

>KTI 3 CAA45778.1 217aa

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLDSGGTYIILSDITAFGGIRAAPTGNERCPLTVVQ
SRNELDKGIGTISSPYRIRFIAEGNPLRLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAVD
GWFLRLERVSDDEFNNYKLVFCTQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKVDKESLAKKN
HGLSRSE

Kunitz Trypsin Inhibitor 3 Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

		1	50
KTI 3	(1)	MKSTIFF-LFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITA	
KTI 3 AAB23464.1	(1)	MKSTIFF-LFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITA	
KTI 3 CAA45777.1	(1)	MKSTIFFALFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITA	
KTI 3 CAA45778.1	(1)	MKSTIFFALFLFCAFTTSYLPSAIADFVLDNEGNPLDSGGTYIILSDITA	
Consensus	(1)	MKSTIFFALFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITA	
		51	100
KTI 3	(50)	FGGIRAAPTGNERCPLTVVQSRNELDKGIGTISSPYRIRFIAEGHPLSL	
KTI 3 AAB23464.1	(50)	FGGIRAAPTGNERCPLTVVQSRNELDKGIGTISSPYRIRFIAEGHPLSL	
KTI 3 CAA45777.1	(51)	FGGIRAAPTGNERCPLTVVQSRNELDKGIGTISSPYRIRFIAEGHPLSL	
KTI 3 CAA45778.1	(51)	FGGIRAAPTGNERCPLTVVQSRNELDKGIGTISSPYRIRFIAEGNPLRL	
Consensus	(51)	FGGIRAAPTGNERCPLTVVQSRNELDKGIGTISSPYRIRFIAEGHPLSL	
		101	150
KTI 3	(100)	KFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFLRLERVSD	
KTI 3 AAB23464.1	(100)	KFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFLRLERVSD	
KTI 3 CAA45777.1	(101)	KFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFLRLERVSD	
KTI 3 CAA45778.1	(101)	KFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAVDGWFLRIERVSD	
Consensus	(101)	KFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGWFLRLERVSD	
		151	200
KTI 3	(150)	DEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQK	
KTI 3 AAB23464.1	(150)	DEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQK	
KTI 3 CAA45777.1	(151)	DEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQK	
KTI 3 CAA45778.1	(151)	DEFNNYKLVFCTQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQK	
Consensus	(151)	DEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQK	
		201	217
KTI 3	(200)	LDKESLAKKNHGLSRSE	
KTI 3 AAB23464.1	(200)	LDKESLAKKNHGLSRSE	
KTI 3 CAA45777.1	(201)	LDKESLAKKNHGLSRSE	
KTI 3 CAA45778.1	(201)	VDKESLAKKNHGLSRSE	
Consensus	(201)	LDKESLAKKNHGLSRSE	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

>KTI 3 216aa

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITAFGGIRAAPTGNERCPLTVVQS
RNELDKGIGTISSPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDG
WFLRLERVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDKESLAKKNH
GLSRSE

Supporting Information S14: Gly m 8 Consensus Sequence Generation

Gly m 8 Protein Sequences:

>Gly m 8 NP_001238443 158aa
MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQLQGVNLTPEKHIMEKIQGRGDDDDDDDDNNHILRT
MRGRINYIRRNEGKDEDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEKELINL
ATMCRFGPMIQCDLSSDD

>Gly m 8 Glyma13g36400.1 174aa
MPPPSLHFTSPINSKMTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQLQGVNLTPEKHIMEKIQGRG
DDDDDDDDNNHILRTMRGRINYIRRNEGKDEDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEL
EEKQKKKMEKELINLATMCRFGPMIQCDLSSDD*

>Gly m 8 NP_001234950 158aa
MTKFTILLISLLFCIAHTCSASEWQHQQDSCRKQLQGVNLTPEKHIMEKIQGRGDDDDDDDDNNHILRT
MRGRINYIRRNEGKDEDEEEEGHMQKCRTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEKELINL
ATMCRFGPMIQCDLSSDD

>Gly m 8 Glyma12g34160.1 156aa
MTKLTILLIALLFIAHTCCASKWQHQHQQESCREQLKGINLNPCEHIMEKIQAGRRGEDGSDHILIRTM
PGRINYIRKKEGKEEEEGHMQKCCSEMSELKSPICQCKALQKIMDNQSEQLGKEKKQMERELMNLAIR
CRLGPMIGCDLSSDD*

Gly m 8 Protein Sequence Alignments: (highlighting colors indicate shared peptide sequences among available proteins)

Gly m 8 NP_001238443	(1)	-----MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQL	50
Gly m 8 Glyma13g36400.1	(1)	MPPPSLHFTSPINSKMTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQL	
Gly m 8 NP_001234950	(1)	-----MTKFTILLISLLFCIAHTCSASEWQHQQDSCRKQL	
Gly m 8 Glyma12g34160.1	(1)	-----MTKLTILLIALLFIAHTCCASKWQHQHQQESCREQL	
Consensus	(1)	MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQL	
Gly m 8 NP_001238443	(36)	QGVNLTPEKHIMEKIQGRGDDDDDDNNHILRTMRGRINYIRRNEGKD	100
Gly m 8 Glyma13g36400.1	(51)	QGVNLTPEKHIMEKIQGRGDDDDDDNNHILRTMRGRINYIRRNEGKD	
Gly m 8 NP_001234950	(36)	QGVNLTPEKHIMEKIQGRGDDDDDDNNHILRTMRGRINYIRRNEGKD	
Gly m 8 Glyma12g34160.1	(36)	KGINLNPCEHIMEKIQAGRRGEDGSDHILIRTMPGRINYIRKKEGKE	
Consensus	(51)	QGVNLTPEKHIMEKIQGRGDDDDDDNNHILRTMRGRINYIRRNEGKD	
Gly m 8 NP_001238443	(86)	EDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEK	150
Gly m 8 Glyma13g36400.1	(101)	EDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEK	
Gly m 8 NP_001234950	(86)	EDEEEEGHMQKCRTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEK	
Gly m 8 Glyma12g34160.1	(85)	EEE--EGHMQKCCSEMSELKSPICQCKALQKIMDNQSEQLGKEKKQMER	
Consensus	(101)	EDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEK	
Gly m 8 NP_001238443	(136)	ELINLATMCRFGPMIQCDLSSDD-	174
Gly m 8 Glyma13g36400.1	(151)	ELINLATMCRFGPMIQCDLSSDD-	
Gly m 8 NP_001234950	(136)	ELINLATMCRFGPMIQCDLSSDD-	
Gly m 8 Glyma12g34160.1	(133)	ELMNLAIRCRLGPMIGCDLSSDD-	
Consensus	(151)	ELINLATMCRFGPMIQCDLSSDD	

Based on the above sequence alignments and overall sequence homology among the available sequences, the following sequence was used for method development. This sequence was selected based on degree of conservation and homology with all other sequences available.

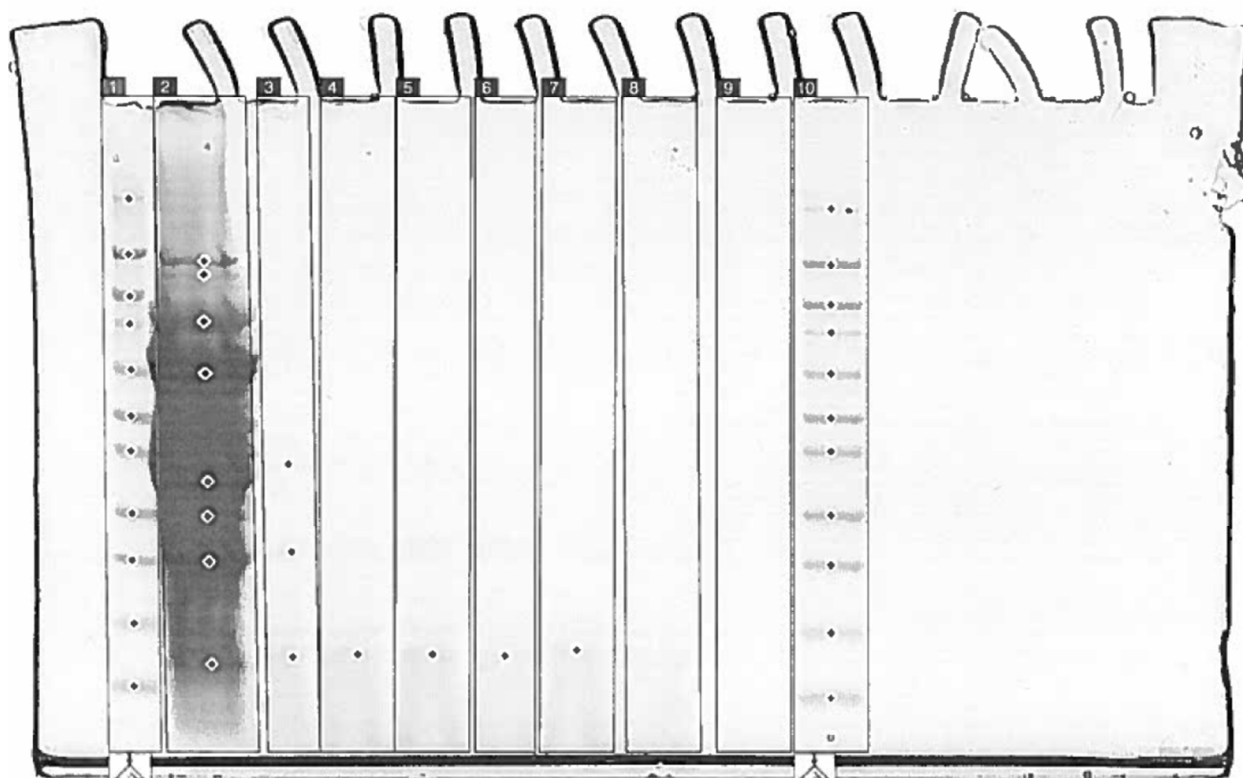
>Gly m 8 NP_001238443 158aa
MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQLQGVNLTPEKHIMEKIQGRGDDDDDDDDNNHILRT
MRGRINYIRRNEGKDEDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEEEEEKQKKKMEKELINL
ATMCRFGPMIQCDLSSDD

Supporting Information S14: Confirmatory SDS-PAGE gel for digestion completeness

30 μ L of digested sample (for individual time points) and 10 μ L of 4X Laemmli sample buffer with 5 % (v/v) β -mercaptoethanol were added to 2 mL low binding tubes. Samples were heated at 95 $^{\circ}$ C for 5 minutes and centrifuged for 2 minutes at > 20,000g. The total samples and molecular weight markers were loaded onto the gel and run for 130 min at 100 volts. The gel was transferred to a container containing 100 mL deionized water and gently shaken on an orbital shaker for 15 min. The gel was rinsed 3-times with 100 mL deionized water. 50 mL ThermoScientific blue coomassie stain (cat #24590) was added to the gel and shaken overnight. Gel was equilibrated with deionized water for 1 hour followed by addition of drying solution (Bio Rad catalogue number 161-0752) and shaken for 30 minutes.

Gel Loading Pattern:

Lane m	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7	Lane 8	Lane 9	Lane 10	Lane 11	Lane 12	Lane M
marker	0 30uL	4 30uL	8 30uL	10 30uL	12 30uL	15-1 30uL	15-2 70 of 80 uL	N/A	marker	N/A	N/A	N/A	N/A



Supporting Information S14: Confirmatory SDS-PAGE gel for digestion completeness

