Proteomic analysis of anti-nutritional factors (ANF's) in soybean seeds as affected by environmental and genetic factors

K.M. Maria John^a, Farooq Khan^b, Davanand L Luthria^a, Wesley M Garrett^c and Savithiry Natarajan^{d*}

^aUSDA-ARS, Beltsville Human Nutrition Research Center, Beltsville, Maryland 20705, United States

^bUniversity of Maryland, Department of Plant Science & Landscape Architecture, College Park, MD 20742, USA

^cUSDA-ARS, Animal Biosciences and Biotechnology Laboratory, Beltsville, MD 20705, USA

^dUSDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA

Running Title: ANF's of soybean affected by environmental and genetic factors

***Corresponding Author**

Dr. Savithiry Natarajan, Ph.D.

Soybean Genomics and Improvement Laboratory, ARS-USDA

10300, Baltimore Avenue, Beltsville, MD 20705

Phone Number: 301-504-5258

Fax Number: 301-504-5728

Email: savi.natarajan@ars.usda.gov

ABSTRACT

The genotype (G), environment (E), and the relationship between G and E on soybean seed anti-nutritional factors (ANFs) were examined under three different agro-climatic conditions. The field trials were conducted at Maryland, South Carolina and South Dakota using nine region specific genotypes. At each location, the nine genotypes were grown with two planting/sowing dates. Differentially expressed protein spots from the two-dimensional gel electrophoresis were analyzed using mass spectrometry. Seven ANFs corresponding to soybean agglutinin and Kunitz trypsin inhibitor were identified based on the statistical significance levels at $p < 0.005$. The G and environmental conditions (planting/sowing season) influences the ANF's content. This initial study suggests that early sowing reduces the total ANF's content irrespective of genotypes and their growing locations.

Keywords: Anti-nutritional factor, Environment, Genotypes, Soybean, 2D-PAGE, MALDI-TOF-MS, Proteomics.

1. Introduction

Soybean is considered to be among one of the major economic crops due to the oil and in-expensive vegetable protein that can be resourced from it, which is valuable to humans and animals. The soybean proteins are widely used in various processed foods throughout the industrialized world. Soybean has also been reported for beneficial effects, such as reduction in the risk of cancer, diabetes, obesity and cardiovascular diseases (Cederroth, Zimmermann & Nef, 2012; Chen, Erh, Su, Liu, Chou & Cheng, 2012; Nemitz, Moraes, Koester, Bassani, von Poser & Teixeira, 2015). Due to the above beneficial effects, the global demand for the soybean and their processed food is increasing. However, soybean is listed at the big 8 among the allergic diet for humans and animals due to the presence of many kinds of allergens and anti-nutritional factors (ANF's) (Grant, 1989; Liener, 1994). According to food allergens report, about 2% of adults and 5-8% of children have allergies due to consumption of soybean and the soy-derived food products.

Soybean agglutinin (SBA) and Kunitz trypsin inhibitor (KTI) are the largest ANFs in the soybean diet. These proteins are primarily protease inhibitors. KTI inhibits trypsin, an important animal digestive enzyme and also involved in respiratory hypersensitivity reactions (Besler, Helm & Ogawa, 2000). According to Jofuku & Goldberg (1989), the KT1 and KT2 have identical nucleotide sequences, while the KT3 gene has diverged about 20% and this protein contains 32% sequence homology with a rye grass pollen allergen (Burks, Cockrell, Connaughton, Guin, Allen & Helm, 1994). The SBA is an oligomeric β -sheet protein, found in the family of legume lectins. These carbohydrate binding proteins are involved in various biological recognition processes (Lis & Sharon 1998). This protein makes changes in the integrity and permeability of the epithelial layer and hence allows small molecules to cross the

epithelial barrier (Dalla Pellegrina, Rizzi, Mosconi, Zoccatelli, Peruffo & Chignola, 2005). The processing of soybean based products alters the ANF's levels when dry-or moist-heated for several hours (Samadi & Yu, 2011). Continuous heat exposure denature the protein to varying degrees from its tetrameric biologically-active form into monomers or remaining in part as a tetramer. Genotypic, cultivation, and physiological status of the soybean possess variation in their ANF's levels have been reported in previous studies (Ritt, Mulinari, Vasconcelos & Carlini, 2004; Peumans & van Damme, 1995; Etzler, 1996).

Soybean is the crop containing many genotypic and phenotypic variations based on their environmental and cultivation conditions resulting in variations of proteins, carbohydrates, lipids and amino acids. USA is among one of the major producers and exporters of soybean and their products. Soybeans were cultivated in different parts of the USA where their agro-climatic conditions are extremely different from one region to another. Not all the soybean genotypes respond well under same location and hence the 27 genotypes were selected in this study to represent a wide range of diversity (wild type, Asian landraces and cultivars developed in US) and optimally grow in that location. A previous report related to soybean proteomics on seed composition, specifically, allergens and storage protein affected by genotypes and environmental conditions clearly indicates that the 47 storage and 8 allergen proteins were statistically significant differences in expression among environmental conditions compared to genotype and environmental interaction (Natarajan et al., 2016). However, the effect of different sowing season on the ANFs and their interaction between genotypes and environment were not reported. The changes in the sowing season provide difference in the seed germination environmental conditions and hence their effect on ANF's might be an interesting aspect for the soybean based product development. Natarajan (2010) characterized soybean proteins using 2-D PAGE

followed by MALDI-TOF-MS. The storage and allergen protein expression and metabolic pattern differences among the genotypes and growing locations are reported by Natarajan et al. (2016) and Maria John, Natarajan & Luthria (2016).

The objective of this investigation was to determine the variation of ANF's in the 27 different genotypes grown under two different sowing time period. This experiment was conducted in three different agro-climatic conditions in Maryland (MD), South Carolina (SC), and South Dakota (SD) of USA.

2. Materials and methods

2.1. Chemicals

All the chemicals and reagents were purchased from GE Healthcare (Piscataway, NJ, USA) and Sigma-Aldrich Chemical Co. (St. Louis, MO, USA). Immobilized pH gradient (IPG) strips, urea, ampholytes (pH 4.0-7.0), acrylamide, bis-acrylamide, SDS, TEMED, ammonium persulfate, thiourea, dithiothreitol, and CHAPS were from GE Healthcare. Tris-HCl (pH 8.8), 2-mercaptoethanol (2-ME), trichloroacetic acid (TCA) and glycerol were purchased from Sigma-Aldrich Chemical Co (Natarajan et al., 2016).

2.2. Plant materials

The soybean genotypes were obtained from the USDA Soybean Germplasm Collection, Urbana, IL. Three planting locations, such as Maryland (MD), South Carolina (SC) and South Dakota (SD), were chosen for this experiment. Nine different region specific soybean genotypes (G), such as PI 407027, PI 562565 and PI 479752 (*Glycine soja*); PI 399043, PI 567298, and PI 594615 (*Glycine max*); Williams 82, Manokin, and Essex (standard check genotypes) were

cultivated in Beltsville, MD. The genotypes PI 407140, PI 407282 and PI 464935 (*G. soja*); PI 097094, PI 587799, and PI 588053A (*G. max*); Dillon, Benning, and Motte (standard check genotypes) were cultivated in Clemson, SC. The genotypes PI 407288, PI 458536 and PI 326582A (*G. soja*); PI 81775, PI 423954, and PI 603318 (*G. max*); Surge, Lambert, and Parker (standard check genotypes) were cultivated at Brookings, SD (Table 1).

2.3. Field experiments

The soybeans were cultivated using a randomized complete block design on hillplots (spacing: 91 cm between hills, 61 to 76 cm between rows) with five seeds per hill, thinning to two plants per hill. Chemical and mechanical methods were adapted to control the weeds. At each location, the soybean seeds were planted twice, one month apart and designated as “early” (May, 2005) sowing and “late” (June, 2005) sowing. These two season provides different environmental conditions to the planted seeds growth and development and hence considered as different environmental sample (E). The second planting was in the same field adjacent to the first planting. Other field and planting conditions were identical to the first planting. In some instances (MD late genotype PI 562565 and SC early genotype PI 97094), we did not obtain sufficient seed from all three replicates.

2.4. Extraction of proteins from soybeans

The seeds from each location (MD, SC and SD) sown at two different times (Early and Late) were collected separately and powdered in liquid nitrogen. The powders were used for the protein extraction using TCA–acetone–2-mercaptoethanol for the precipitation of proteins followed by washing with acetone and 2-mercaptoethanol. Lysis buffer (9 M urea, 1% CHAPS,

1% (w/v) ampholytes (pH 4–7), 1% DTT) was used to solubilize the proteins and were centrifuged to remove the insoluble materials. The protein were checked for their concentration using Bradford method, prior to analysis by 2D-PAGE (Natarajan, Xu, Caperna & Garrett, 2005).

2.5. 2D-PAGE analysis of ANF's

The first dimension isoelectric focussing (IEF) was performed in a IPGphor II system (GE Healthcare, Piscataway, NJ) using 13 cm pH 4.0-7.0 linear IPG strips with 100 µg protein (Natarajan, Xu, Caperna & Garrett, 2005). For the second dimension, the IPG strips were incubated with 50mM Tris-HCl pH 8.8, 6M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, and 1% DTT for 20 min, acetylated with iodoacetamide and subsequently placed onto 12% polyacrylamide gels as described by Natarajan et al., 2005. A Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ) was used according to the manufacturer's recommendations for the second dimension gel. The 2D-PAGE gels were stained with Colloidal Coomassie Blue G-250 according to Natarajan, Xu, Caperna & Garrett, (2005). The gels were stored in 20% ammonium sulfate solution and scanned using laser densitometry (GE Healthcare, Piscataway, NJ). Triplicate samples of the seeds obtained from each hillplot were used both for protein extractions and 2D-PAGE analysis.

2.6. Protein expression analysis

The protein intensities were analysed using Progenesis SameSpots (TotalLab, Newcastle, England) software and the images of gels were created through scans in the format of Maya Embedded Language (MEL). The gel images were aligned and the spots were detected

152 automatically. In addition, a manual inspection was made to remove the non-spot fragments
153 falsely reported by the software. Intensity of the spots were presented in terms of spot area and
154 volume. All of the differentially expressed protein spots were then filtered to only include those
155 with proper intensity replication (p -value <0.05) using the software-based ANOVA.

157 2.7. *In-gel protein digestion and mass spectrometry*

158 The protein spots were manually removed from the stained gel followed by de-stained,
159 washed, and dehydrated with acetonitrile. Modified porcine trypsin (Promega, Madison, WI) in
160 ammonium bicarbonate buffer was used to digest (overnight at 70 °C) the protein. The resulting
161 peptides were extracted, dried, and dissolved in 50% acetonitrile and 0.1% trifluoroacetic acid
162 (Natarajan, Xu, Caperna & Garrett, 2005). MALDI-TOF-MS and LC-MS/MS were used for the
163 identification of proteins. Dried droplet method with α -cyanohydroxycinnamic acid (CHCA)
164 matrix was employed for the protein samples to analyze with MALDI-TOF-MS (Applied
165 Biosystems, Framing- ham, MA) using trypsin at 842.51 and 2,211.10 m/z as internal standards
166 for the spectral calibration. Thermo Finnigan LCQ Deca XP plus Ion Trap LC-MS system was
167 used to analyze proteins that were not positively identified by MALDI-TOF-MS. Acetonitrile
168 and water containing 0.1% formic acid with the gradient begins from 5% to 60% for a period of
169 30 min on a reverse phase column served as mobile phase. The instrument was operated with a
170 duty cycle that acquires MS/MS spectra on the three most abundant ions identified by a survey
171 scan from 300 to 2000 Da. Dynamic exclusion was employed to prevent the continuous analysis
172 of the same ions. The raw data were processed by Sequest to generate DTA files for database
173 searching. The merge pl script from Matrix Science was used to convert multiple Sequest DTA
174 files into a single mascot generic file suitable for searching in Mascot.

2.8. Data interpretation using databases

The peptides resulting from trypsin digestion were subjected to MALDI-TOF/MS analysis to begin the process of determining their molecular mass. These resultant mass to charge ratio (m/z) values from the MALDI-TOF analysis were used to search theoretically generated peptide masses from known protein sequences. Protein identification was performed by searching the National Center for Biotechnology Information (NCBI) non-redundant database using the public Mascot search engine (www.matrixscience.com) that uses a probability-based scoring system. The following parameters were used for database searches with MALDI-TOF-MS peptide mass fingerprinting data: monoisotopic mass, 25 ppm mass accuracy, trypsin as digesting enzyme with one missed cleavage allowed, and carbamidomethylation of cysteine, oxidation of methionine, N-terminal pyroglutamic acid from glutamic acid or glutamine as the allowable variable modifications. To qualify the MALDI-TOF-MS data as a positive identification, a protein's score had to be equal to or exceed the minimum significant score. Selective samples that did not result in positive identification through MALDI-TOF-MS were subsequently subjected to LC-MS/MS analysis. For database searches with MS/MS spectra, the following parameters were used: average mass, 1.0 Da peptide and MS/MS mass tolerance, peptide charge of +1, +2, or +3, trypsin as digesting enzyme with one missed cleavage allowed, carbamidomethylation of cysteine as a fixed modification, oxidation of methionine, N-terminal pyroglutamic acid from glutamic acid or glutamine as allowable variable modifications. For LC-MS/MS, the positive identification required a minimum of two unique peptides, with a significant ion score in at least one of the peptides.

2.9. Multivariate and statistical analysis of data

A targeted metabolic analysis was performed using the data sets obtained from the protein intensity corresponding to different genotypes, sowing seasons, and cultivation regions. The data sets were processed using Partial least square discriminant analysis leads (PLS-DA) model (SIMCA 13.0.3, UMETRICS, Umea, Sweden). For the statistical analysis, STATISTICA Version 7 (StatSoft, Tulsa, OK, USA) software was used to determine the presence of significant differences among genotypes grown at each location, among environments and genotype $\times$ environment interaction using analysis of variance. The source of variation was partitioned into genotypes (G), environments (sown time (E), and interaction between genotype and environments (G $\times$ E).

3. Results and discussion

3.1. General

In this study, the environmental (E) and genotypic (G) variations of ANF's profile in early and late sown soybean cultivated in three distinct geographical locations, namely MD, SD, and SC, were examined. At each location, nine different soybean genotypes (optimally grown at particular location) comprising wild soybean, Asian landraces, and commercial soybean cultivars were selected (Table 1). Physiological and morphological examination of the field grown plants were recorded for comparison purposes (data not presented).

The 2D-PAGE gel image representing the ANF's profile with reference to MD genotype (Williams 82) expressed based on their environmental (early sown) conditions were presented in Fig. 1. A total of seven ANF's were identified and labeled spot #1 to 7 in the gel. Identification of the ANF's was achieved using MALDI-TOF-MS and LC-MS/MS. The theoretical values for

isoelectric point, molecular weight, NCBI accession number; number of unique peptides matches, percentage of parent protein sequence covered and MS Score were considered for the characterization and identification of the ANFs. Soybean agglutinin (spots #1 to 5) and Kunitz trypsin inhibitor (spots #6 and 7) and their identification particulars were presented in Table 2. For each genotypes (G) and environmental (E) conditions, three gels were used for the separation and analysis of the proteins.

3.2. Genotypic variations (G) of ANF's analyzed by PLS-DA model

The genotypic variations of ANF's content from 27 genotypes grown under three different agro-climatic conditions were examined and reported. The PLS-DA based clustering model was performed to visualize the differences between the genotypes based on their ANF's content. The MD grown soybean genotypes are presented in Fig. 2a (early sown) and 2b (late sown). Total of 57.1% and 58.3% of variations were captured by the two components for early and late sown soybean, respectively. PI407027, Williams 82 and Manokin genotypes clustered together in the case of early sown season (Fig. 2a) whereas PI567298, PI594615, Williams 82, Manokin and Essex were clustered in a single group with late sown season (Fig. 2b). In case of SC grown soybean genotypes, three distinct clusters were observed with early sown condition, PI407140 and PI097094 genotypes clustered as a single group, PI407282 was distinctly separated from other six genotypes which clustered as single group (Fig. 2c). Among the late sown genotypes, PI407282 clustered separately as a single group. PI407140 and PI464935 showed close similarities were grouped together and the other six genotypes did not show clear differentiation were pooled as one subgroup. The PLS model captures 72.1% of variations between the genotypes (Fig. 2d). The SD genotypes analyzed by PLS model captured 54.6%

(early sown) and 51.6% (late sown) variations in two components (Fig. 2e and f). The two genotypes, Lambert and Parker showed close similarities in both sowing seasons (early and late). Comparatively high plant to plant variations were observed in SD grown soybean genotypes.

3.3. Environmental (E) variations of ANF's analyzed by PLS-DA model

A targeted metabolic approach was made to visualize the differences between the ANF's expression levels between the genotypes (G) and their environmental (E) factors (early vs late sown) and was compared using PLS-DA models (Fig. 3). Results revealed that the response of protein expression between the genotypes significantly varied between growing conditions and seasons with MD and SC grown soybean genotypes. The MD genotypes showed 47.4% (PLS-1) and 17.7% (PLS-2) variation between the growing seasons (Fig. 3a). The SC (Fig. 3b) genotypes captures maximum levels of variation between the growing seasons, such as 75.8% (PLS-1) and 9.62% (PLS-2). In case of SD grown genotypes, they showed two groups in general but some of the genotypes, such as Lambert and Parker, showed a mixed variation between the growing seasons (Fig. 3c). Variations between the samples were captured with 49.1% (PLS-1) and 11.7% (PLS-2). This evidently proves that the ANF's expression pattern differs with growing locations and seasons. The SD genotypes grown under late condition showed very high variation between their plant to plant variations. All the genotypes grown in MD, SC and SD showed high variation in their expression levels based on their environmental conditions.

3.4. ANF's expression pattern response to environment

The genotypes and their response in ANF's expression during both early and late sown types are presented in Table 3. Among the MD grown genotypes, most of the Soybean agglutinin

intensities were found low in PI562565 genotype from early sowing season. However, the late sown of this genotype does not produce enough seeds for protein analysis and hence the samples were not compared. Soybean agglutinin (spot 2, 4) and kunitz trypsin inhibitors (spot 6 and 7) registered lower intensities expression with PI479752 genotypes. In all the genotypes the early sowing registered lower intensities of the ANF's. Among the three agro-climatic conditions, the SC grown genotypes contains lower levels of total ANF's content was observed. The MD genotypes resulted in variation between ~9 to ~55% of ANF's expression. As like MD genotypes, the SC (~0.5 to ~23%) and SD (~28 to ~59%) genotypes also responded to low intensities of ANF's in the early sowing compared with late sowing season.

The genotypic and environmental variations on seed protein content of soybean has been well documented by Maestri, Labuckas, Meriles, Lamarque, Zygadlo & Guzman, (1998); Piper & Boote, (1999); Natarajan, Xu, Bae, Caperna & Garrett, (2006); Bellaloui, Yin, Mengistu, McClure, Tyler & Reddy, (2013). However, reports on the response of ANF's levels due to environment and genotypic conditions were limited. Improving nutritive values of the soybean depends upon the content of ANF's. Reducing the ANF's content in the soybean and soybean based food products requires additional processing steps. Production of transgenic, microbial fermentation and thermal processing on ANF's content are well documented. For an example, the trypsin inhibitor and protease inhibitor were reduced from 1.20 to 0.010 mg/g when the soybeans are treated by *Lactobacillus* sps (Adeyemo & Onilude 2013). In a separate study, Teng, Gao, Yang, Liu, Tian & Wang, (2012) used solid-state fermentation using the *Aspergillus oryzae* or *Bacillus subtilis* microbes reduce trypsin inhibitor contents by 96% and 82%, respectively. In the case of *Saccharomyces cerevisiae*, decomposed carbohydrate-based but not protein-based ANF's was reported by Chi & Cho (2016).

Thermal processing, such as cooking and roasting on the decomposition of ANF's content was reported by Adeyemo & Onilude (2013). Heat is an important inactivation factor of ANF's and hence extrusion, fluidized bed, spouted bed, and infrared radiation were investigated with reference to moisture reduction, urease inactivation, protein solubility and lysis. Report showed that these techniques reduced the urease activity and hence the remaining lysine being insignificantly different under the testing conditions (Wiriyapaiwong, Soponronnarit & Prachayawarakorn, 2004). Electron beam irradiation was reported to degrade ANF's in cottonseed, soybean and canola seeds. Trypsin inhibitor activity of 15, 30 and 45 kGy EB-irradiated soybean was decreased by 19, 73 and 88%, respectively and it was concluded that EB-irradiation was an effective processing method for improving the nutritive value of soybean (Ebrahimi-Mahmoudabad & Taghinejad-Roudbaneh, 2011).

Recently production of transgenic soybean against their reduced biosynthesis of ANF's was reported by gene silencing and mutant development. Krishnan (2001) reported that the soybean mutant PI 196168 reduced Kunitz trypsin inhibitor levels during seed development. However, PI 196168 contained only four zones of enzyme inhibition. Lower amounts of KTi3 mRNA when compared with Amsoy 71 were reported based on northern blot analysis. Fracasso, Perussello, Haminiuk, Jorge & Jorge, (2014) compared two transgenic soybeans (A7321 and CD231) with conventional genotypes (CD206 and BRS232) grown under five different temperatures (25, 35, 45, 55 and 65 °C). The temperature showed strong dependence on their reduced ANF's content in transgenic and conventional cultivars. The differences were not clear towards hydration rate or equilibrium moisture content.

Above literatures showed that the processing (microbial, thermal, and hydration) or mutant genotypes showed reduced content of ANF's. Effects of environment and interactions

between the environment and genotypes on allergens of soybean were well documented by Stevenson, Woods, Hong, Kong, Thelen & Ladics, (2012) and Natarajan et al. (2016). Different agro-climatic conditions showed significant changes of allergen proteins levels when compared to genotypic protein variation. In addition, the levels of β -conglycinin/glycinin were highly affected by the location compared to their genotypic variations (Fehr et al., 2003; Mujoo, Trinh & Ng, 2003; Natarajan, Xu, Bae, Caperna & Garrett, 2006). However, only limited studies were discussed about the ANF's content on genotypes as well as environmental conditions. The genotypic variations and heat treatment on ANF's were reported by Shivakumar et al. (2015) and Brune, Pinto, Peluzio, Moreira & de Barros, (2010) stated that heat treatment reduces the accumulation of ANF's. Comparing the late sown season, the early sown seeds developed and mature during the time of hot weather might interfere the ANF's levels. Perhaps the heat stress destruct or translate several essential amino acids like cystine and lysine unavailable during seed maturation. Study of the level of ANF's in the genotypes will be helpful for the improved soybean varieties development through breeding strategies. Apparently developing mutants or new genotypes might be helpful for certain extent only because of their physiological changes and its effect on nutritive and quality functionalities of soybean.

These results suggest that the sowing time significantly decreases the biosynthesis of ANF's. In all the three agro-climatic condition same trend of reduced AFN's was observed in early sown soybean. Even though the phenotypic variations are present in the biosynthesis of ANF's content, the environmental changes highly influenced their biosynthesis. ANF's were reduced by heat was already reported as one of the nutritional development of soy based products. Early sowing soybean set seeds at the summer time which may be a cause of the physiological reduction in ANF's content of the soybean. Polymorphism of KTI genes and their

genotypic variations among different soybean genotypes were reported by Natarajan, Xu, Bae & Ba, (2007).

4. Conclusions

Soybean plays an important role in animal feed and food production as it provides a major source of vegetable protein. However, it is documented in literature that soybean contains various kinds of ANF's that affects the nutritional value, utilization, and digestibility of soybean protein. Industrialized soybean products include additional processing steps to decomposing the ANF's content. Alternatively, transgenic based approaches have been used to reduce ANF's content. In the present study, 27 soybean genotypes grown at three different locations under two different environmental (sowing season) conditions were compared for their ANF's content. Results revealed that all the genotypes grown at three locations drastically reduced in their total ANF's content was observed with early sown season.

Acknowledgements

The authors wish to thank Dr. J. Slovin for her critical review of this manuscript. The authors wish to thank Dr. E. Shipe and Dr. R. Scott for providing soybean seeds. Funding for this research was provided by ARS project 1245-21220-232-00D. Mention of trade name, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

References

359 Adeyemo, S.M., & Onilude A.A. (2013). Enzymatic reduction of anti-nutritional factors in
360 fermenting soybeans by *Lactobacillus plantarum* isolates from fermenting cereals.
361 *Nigerian Food Journal*, 31, 84-90

362 Bellaloui, N., Yin, X., Mengistu, A., McClure, A.M., Tyler, D.D., & Reddy, K.N. (2013).
363 Soybean seed protein, oil, fatty acids, and isoflavones altered by potassium fertilizer rates
364 in the midsouth. *American Journal of Plant Sciences*, 4, 976 – 988

365 Besler, M., Helm, R.M., & Ogawa, T. (2000). Allergen data collection - update: Soybean
366 (*Glycine max*). *Internet Symposium On Food Allergens*, 2, 1-35.

367 Brune, M.F.S.S., Pinto, M.S., Peluzio, M.C.G., Moreira, M.A., & de Barros, E.G. (2010).
368 Biochemical and nutritional evaluation of a soybean line lacking the Kunitz trypsin
369 inhibitor and lectins. *Ciencia e Tecnologia de Alimentos*, 30, 657-663

370 Burks, A.W., Cockrell, G., Connaughton, C., Guin, J., Allen, W., & Helm, R.M. (1994).
371 Identification of peanut agglutinin and soybean trypsin inhibitor as minor legume
372 allergens. *International Archives of Allergy and Immunology*, 105, 143–9.

373 Cederroth, C.R., Zimmermann, C., & Nef, S. (2012). Soy phytoestrogens and their impact on
374 reproductive health. *Molecular and Cellular Endocrinology*, 355, 192–200.

375 Chen, K.I., Erh, M.H., Su, N.W., Liu, W.H., Chou, C.C., & Cheng, K.C. (2012). Soy foods and
376 soybean products: from traditional use to modern applications. *Applied Microbiology and*
377 *Biotechnology*, 96, 9–22

378 Chi, C.H., & Cho, S.J. (2016). Improvement of bioactivity of soybean meal by solid-state
379 fermentation with *Bacillus amyloliquefaciens* versus *Lactobacillus* spp. and
380 *Saccharomyces cerevisiae*. *LWT - Food Science and Technology*, 68, 619-625

381 Dalla Pellegrina, C., Rizzi, C., Mosconi, S., Zoccatelli, G., Peruffo, A., & Chignola, R. (2005).
 382 Plant lectins as carriers for oral drugs: is wheat germ agglutinin a suitable candidate?.
 383 *Toxicology and Applied Pharmacology*, 207, 170–178

384 Ebrahimi-Mahmoudabad, S.R., & Taghinejad-Roudbaneh, M. (2011). Investigation of electron
 385 beam irradiation effects on anti-nutritional factors, chemical composition and digestion
 386 kinetics of whole cottonseed, soybean and canola seeds. *Radiation Physics and*
 387 *Chemistry*, 80, 1441-1447

388 Etzler, M.E. (1996). In: van Driessche, E., Rouge, P., Beeckmans, S., Bog-Hansen, T.C. (Eds.),
 389 Lectins, Biology, Biochemistry, Clinical Biochemistry, Vol. 11. Hellerup, Denmark, 3-9.

390 Fehr, W.R., Hoeck, J.A., Johnson, S.L., Murphy, P.A., Nott, J.D., Padilla, G.I., & Welke, G.A.
 391 (2003). Genotype and environment influence on protein components of soybean. *Crop*
 392 *Science*, 43, 511–514.

393 Fracasso, A.F., Perussello C.A., Haminiuk C.W.I., Jorge L.M.M., & Jorge R.M.M. (2014).
 394 Hydration kinetics of soybeans: Transgenic and conventional cultivars. *Journal of Cereal*
 395 *Science*, 60, 584-588

396 Grant G. (1989). Anti-nutritional effects of soybean: A review. *Progress in food & nutrition*
 397 *science*, 13, 317–348.

398 Jofuku, K.D. & Goldberg, R.B. (1998). Kunitz trypsin inhibitor genes are differentially
 399 expressed during the soybean life cycle and in transformed tobacco plants. *The plant*
 400 *Cell*, 1, 1079-1093.

401 Krishnan, H.B. (2001). Characterization of a soybean (*Glycine max* (L.) Merr.) mutant with
 402 reduced levels of Kunitz trypsin inhibitor. *Plant Science*, 160, 979-986

403 Liener, I.E. (1994). Implications of anti-nutritional components in soybean foods. *Critical*
404 *Reviews in Food Science and Nutrition*, 34, 31–67.

405 Lis, H., & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular
406 recognition. *Chemical Reviews*, 98, 637 -674

407 Maestri, D.M., Labuckas, D.O., Meriles, J.M., Lamarque, A.L., Zygadlo, J.A., & Guzman, C.A.
408 (1998). Seed composition of soybean cultivars evaluated in different environmental
409 regions. *Journal of the Science of Food and Agriculture*, 77, 494 – 498

410 Maria John, K.M., Natarajan, S. & Luthria, D.L. (2016). Metabolite changes in nine different
411 soybean varieties grown under field and greenhouse conditions. *Food Chemistry*, 211,
412 347-355.

413 Mujoo, R., Trinh, D.T., & Ng, P.K.W. (2003). Characterization of storage proteins in different
414 soybean varieties and their relationship to tofu yield and texture. *Food Chemistry*, 82,
415 265–273.

416 Natarajan, S., Xu, C.P., Caperna, T.J., & Garrett, W.A. (2005) Comparison of protein
417 solubilization methods suitable for proteomic analysis of soybean seed proteins.
418 *Analytical Biochemistry*, 342, 214–220.

419 Natarajan, S.S., Xu, C., Bae, H., Caperna, T.J., & Garrett, W.M. (2006). Characterization of
420 storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using
421 proteomic analysis. *Journal of Agricultural and Food Chemistry*, 54, 3114 – 20.

422 Natarajan, S., Xu, C., Bae, H., & Bailey, B.A. (2007). Proteomic and genomic characterization
423 of Kunitz trypsin inhibitors in wild and cultivated soybean genotypes. *Journal of Plant*
424 *Physiology*, 164, 756—763

425 Natarajan, S.S. (2010). Natural variability in abundance of prevalent soybean proteins.
 426 *Regulatory Toxicology and Pharmacology*, 58, S26-S29

427 Natarajan, S., Khan, F., Song, Q., Lakshman, S., Cregan, P., Scott, R., Shipe, E., & Garrett W.
 428 (2016). Characterization of soybean storage and allergen proteins affected by
 429 environmental and genetic factors. *Journal of Agricultural and Food Chemistry*, 64,
 430 1433-45

431 Nemitz, M.C., Moraes, R.C., Koester, L.S., Bassani, V.L., von Poser, G.L., & Teixeira, H.F.
 432 (2015). Bioactive soy isoflavones: extraction and purification procedures, potential
 433 dermal use and nanotechnology-based delivery systems. *Phytochemistry Reviews*, 14,
 434 849-869

435 Peumans, W.J., & van Damme, E.J.M. (1995). The role of lectins in plant defense.
 436 *Histochemical Journal*, 27, 253-271.

437 Piper, E.L., & Boote, K.J. (1999). Temperature and cultivar effects on soybean seed oil and
 438 protein concentrations. *Journal of the American Oil Chemists' Society*, 76, 1233 –1241.

439 Ritt, A.B., Mulinari, F., Vasconcelos, I.M., & Carlini, C.R. (2004). Antinutritional and/or toxic
 440 factors in soybean (*Glycine max* (L) Merrill) seeds: comparison of different cultivars
 441 adapted to the southern region of Brazil. *Journal of the Science of Food and Agriculture*,
 442 84, 263-270.

443 Samadi., & Yu P. (2011). Dry and moist heating-induced changes in protein molecular structure,
 444 protein sub fraction, and nutrient profiles in soybeans. *Journal of Dairy Science*, 94,
 445 6092-6102.

446 Shivakumar, M., Verma, K., Talukdar, A., Srivastava, N., Lal, S.K., Sapra, R.L., & Singh K.P.
 447 (2015). Genetic variability and effect of heat treatment on trypsin inhibitor content in
 448 soybean (*Glycine max* (L.) Merrill.). *Legume Research*, 38, 60-65
 449 Stevenson, S.E., Woods, C.A., Hong, B., Kong, X., Thelen, J.J., & Ladics, G.S. (2012).
 450 Environmental effects on allergen levels in commercially grown non-genetically
 451 modified soybeans: assessing variation across North America. *Frontiers in Plant Science*,
 452 3, 196
 453 Teng, D., Gao, M., Yang, Y., Liu, B., Tian, Z., & Wang, J. (2012). Bio-modification of soybean
 454 meal with *Bacillus subtilis* or *Aspergillus oryzae*. *Biocatalysis and Agricultural*
 455 *Biotechnology*, 1, 32-38
 456 Wiriyaumpaiwong, S., Soponronnarit, S., & Prachayawarakorn, S. (2004). Comparative study of
 457 heating processes for full-fat soybeans. *Journal of Food Engineering*, 65, 371-382
 458

459 **Table 1. Cultivars of soybean genotypes analyzed for ANF's content grown under three**
 460 **different locations**

S.No	Genotypes	Maturity group	Country of origin	Classification	Experimental location
1	PI 407027	V	Japan	<i>Glycine soja</i>	Maryland
2	PI 562565	IV	South Korea	<i>Glycine soja</i>	Maryland
3	PI 479752	I	China	<i>Glycine soja</i>	Maryland
4	PI 399043	III	South Korea	<i>Glycine max</i>	Maryland
5	PI 567298	V	China	<i>Glycine max</i>	Maryland
6	PI 594615	VI	China	<i>Glycine max</i>	Maryland
7	Williams 82	III	standard check genotypes	<i>Glycine max</i>	Maryland
8	Manokin	IV	standard check genotypes	<i>Glycine max</i>	Maryland
9	Essex	V	standard check genotypes	<i>Glycine max</i>	Maryland
10	PI 407140	VII	Japan	<i>Glycine soja</i>	South Carolina
11	PI 407282	VI	South Korea	<i>Glycine soja</i>	South Carolina
12	PI 464935	VI	China	<i>Glycine soja</i>	South Carolina
13	PI 097094	VII	North Korea	<i>Glycine max</i>	South Carolina
14	PI 587799	VIII	China	<i>Glycine max</i>	South Carolina
15	PI 588053A	VI	China	<i>Glycine max</i>	South Carolina
16	Dillon	VI	standard check genotypes	<i>Glycine max</i>	South Carolina
17	Benning	VII	standard check genotypes	<i>Glycine max</i>	South Carolina
18	Motte	VIII	standard check genotypes	<i>Glycine max</i>	South Carolina
19	PI 407288	II	China	<i>Glycine soja</i>	South Dakota
20	PI 458536	0	China	<i>Glycine soja</i>	South Dakota
21	PI 326582A	II	Russia	<i>Glycine soja</i>	South Dakota
22	PI 81775	I	Japan	<i>Glycine max</i>	South Dakota
23	PI 423954	0	Japan	<i>Glycine max</i>	South Dakota
24	PI 603318	0	China	<i>Glycine max</i>	South Dakota
25	Surge	0	standard check genotypes	<i>Glycine max</i>	South Dakota
26	Lambert	0	standard check genotypes	<i>Glycine max</i>	South Dakota
27	Parker	I	standard check genotypes	<i>Glycine max</i>	South Dakota

461

Table 2. Anti-nutritional factors (ANF's) identified by MALDI-TOF-MS and LC-MS/MS

Spot #	pI / Mr (theor)	Protein Identity	Accession #	# PM	% SC	MS Score	e-Value
1	5.15/27555	Soybean agglutinin	6729836	9	44	90	4.40E-04
2	5.65/30909	Soybean agglutinin	282898	7	10	244	
3	5.15/27555	Soybean agglutinin	6729836	6	35	85	1.40E-03
4	5.15/27555	Soybean agglutinin	3114258	9	41	90	1.50E-04
5	5.15/27555	Soybean agglutinin	3114258	7	30	358	
6	4.61/20310	Kunitz trypsin inhibitor	3318877	15	53	147	3.00E-10
7	4.97/22817	Kunitz trypsin inhibitor	125722	8	42	72	1.10E-02

Spot #, spot number referred to in Fig. 1; pI/Mr (theor), theoretical values for isoelectric point and molecular weight; Accession #, NCBI accession number; #PM, the number of unique peptides matched, %SC, percentage of parent protein sequence covered; MS Score, MOWSE Score

Table 3. Anti-nutritional factors (ANF's) intensity data from field grown soybean genotypes grown in Maryland, South Carolina, and South Dakota field (Early & Late sown)

ANF's	Spot #	Soybean Genotypes - Maryland																	
		PI407027		PI562565		PI479752		PI399043		PI567298		PI594615		Williams 82		Manokin		Essex	
		Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
Soybean agglutinin	1	1.03E+06	8.96E+05	3.60E+05	—	5.94E+05	7.59E+05	6.31E+05	5.79E+05	6.64E+05	6.81E+05	7.00E+05	6.73E+05	8.14E+05	7.09E+05	5.66E+05	6.84E+05	5.87E+05	6.09E+05
Soybean agglutinin	2	1.27E+06	2.09E+06	5.17E+05	—	5.28E+05	1.18E+06	7.78E+05	1.04E+06	9.64E+05	1.57E+06	7.86E+05	1.47E+06	1.04E+06	1.58E+06	9.29E+05	1.30E+06	7.58E+05	1.34E+06
Soybean agglutinin	3	4.15E+05	3.52E+05	4.82E+05	—	4.11E+05	3.56E+05	6.17E+05	4.29E+05	4.19E+05	4.76E+05	4.58E+05	5.15E+05	4.03E+05	3.70E+05	3.97E+05	3.81E+05	4.02E+05	4.41E+05
Soybean agglutinin	4	3.83E+05	2.76E+05	1.60E+05	—	2.41E+05	2.68E+05	4.10E+05	2.57E+05	4.17E+05	6.10E+05	3.71E+05	5.39E+05	5.89E+05	6.01E+05	5.46E+05	5.83E+05	3.97E+05	4.85E+05
Soybean agglutinin	5	9.87E+04	2.84E+05	1.07E+05	—	1.09E+05	3.33E+05	1.31E+05	3.21E+05	1.24E+05	3.79E+05	1.01E+05	3.12E+05	1.33E+05	4.46E+05	9.64E+04	4.74E+05	1.29E+05	4.44E+05
Kunitz trypsin inhibitor	6	3.11E+06	6.77E+06	3.58E+06	—	1.75E+06	3.71E+06	2.75E+06	7.00E+06	3.87E+06	8.98E+06	2.93E+06	6.36E+06	3.24E+06	8.27E+06	3.75E+06	7.04E+06	3.23E+06	7.19E+06
Kunitz trypsin inhibitor	7	4.02E+05	1.50E+06	3.62E+05	—	1.58E+05	7.99E+05	2.74E+05	1.38E+06	3.00E+05	1.25E+06	4.21E+05	1.34E+06	4.00E+05	1.75E+06	4.51E+05	1.45E+06	2.95E+05	9.97E+05
ANF's	Spot #	Soybean Genotypes - South Carolina																	
		PI407140		PI407282		PI464935		PI097094		PI587799		PI588053A		Dillon		Benning		Motte	
		Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
Soybean agglutinin	1	7.83E+03	1.07E+05	4.13E+03	1.49E+05	5.00E+03	2.41E+05	—	6.10E+05	1.56E+04	7.50E+05	9.16E+03	4.75E+05	3.02E+03	6.63E+05	5.14E+03	5.95E+05	3.53E+03	5.12E+05
Soybean agglutinin	2	1.33E+05	2.86E+05	3.41E+05	478600	2.06E+05	3.77E+05	—	5.04E+05	3.38E+05	7.26E+05	3.67E+05	3.94E+05	2.25E+05	5.80E+05	2.64E+05	6.38E+05	2.92E+05	7.77E+05
Soybean agglutinin	3	7.26E+04	1.60E+05	1.22E+04	87420	3.92E+04	1.77E+05	—	2.21E+05	7.75E+04	2.80E+05	1.05E+05	1.11E+05	5.45E+04	2.12E+05	5.96E+04	1.82E+05	7.43E+04	2.85E+05
Soybean agglutinin	4	2.69E+04	4.35E+04	1.14E+04	27150	5.92E+04	8.24E+04	—	2.53E+05	1.38E+05	4.28E+05	9.63E+04	2.17E+05	7.47E+04	3.18E+05	1.10E+05	3.39E+05	1.31E+05	3.10E+05

Soybean agglutinin	5	3.20E+04	8.94E+04	3.79E+04	1.59E+05	3.99E+04	8.69E+04	—	2.36E+05	7.13E+04	2.11E+05	5.07E+04	2.66E+05	6.06E+04	3.20E+05	6.47E+04	2.87E+05	5.86E+04	2.18E+05
Kunitz trypsin inhibitor	6	9.34E+05	2.93E+06	1.34E+06	2.64E+06	9.40E+05	3.54E+06	—	3.70E+06	1.52E+06	3.09E+06	1.23E+06	3.61E+06	1.50E+06	4.08E+06	1.29E+06	4.30E+06	1.05E+06	3.86E+06
Kunitz trypsin inhibitor	7	2.44E+05	3.66E+05	3.63E+05	602100	2.08E+05	4.95E+05	—	6.78E+05	4.15E+05	9.14E+05	4.13E+05	8.14E+05	3.75E+05	8.61E+05	3.19E+05	5.10E+05	3.07E+05	5.42E+05
ANF's	Spot #	Soybean Genotypes - South Dakota																	
		PI407288		PI458536		PI326582A		PI081775		PI423954		PI603318		Surge		Lambert		Parker	
		Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late	Early	Late
Soybean agglutinin	1	6.74E+05	1.14E+06	7.97E+05	6.02E+05	6.97E+05	9.23E+05	5.94E+05	7.13E+05	1.06E+06	1.09E+06	5.82E+05	9.77E+05	6.60E+05	7.68E+05	9.60E+05	7.89E+05	7.14E+05	9.83E+05
Soybean agglutinin	2	4.66E+05	8.86E+05	6.94E+05	742000	5.94E+05	8.04E+05	6.03E+05	7.22E+05	4.61E+05	6.98E+05	5.40E+05	7.16E+05	5.18E+05	8.42E+05	8.94E+05	1.23E+06	6.42E+05	8.34E+05
Soybean agglutinin	3	3.07E+05	3.73E+05	3.32E+05	334300	3.14E+05	4.52E+05	2.73E+05	3.88E+05	3.40E+05	3.71E+05	3.51E+05	4.90E+05	2.79E+05	4.74E+05	3.92E+05	3.99E+05	3.06E+05	3.13E+05
Soybean agglutinin	4	2.81E+05	4.38E+05	2.77E+05	289000	2.49E+05	3.36E+05	3.53E+05	3.62E+05	3.48E+05	4.17E+05	3.27E+05	4.65E+05	3.23E+05	4.15E+05	4.63E+05	4.83E+05	3.44E+05	4.23E+05
Soybean agglutinin	5	1.28E+05	3.57E+05	1.17E+05	2.07E+05	1.58E+05	2.45E+05	1.27E+05	2.63E+05	1.37E+05	2.87E+05	1.57E+05	3.73E+05	1.76E+05	4.13E+05	2.25E+05	3.43E+05	1.80E+05	3.80E+05
Kunitz trypsin inhibitor	6	4.89E+06	5.74E+06	5.18E+06	6.01E+06	4.56E+06	5.47E+06	5.31E+06	5.56E+06	4.70E+06	6.16E+06	5.12E+06	6.26E+06	5.10E+06	6.16E+06	6.42E+06	6.63E+06	6.40E+06	7.75E+06
Kunitz trypsin inhibitor	7	7.58E+05	9.34E+05	7.00E+05	698600	7.40E+05	8.10E+05	5.74E+05	7.78E+05	6.64E+05	7.28E+05	8.06E+05	1.05E+06	7.44E+05	9.30E+05	8.89E+05	1.08E+06	8.59E+05	1.03E+06

" - " Denotes lack of proteomic data due to poor seed yield

Fig. 1. Anti-nutritional factors (ANF's) of Williams 82 (MD – early sown) analyzed by 2-D gel electrophoresis. The pH gradient from 4.0 to 7.0 was used for the first dimension. The second dimension was a 12% SDS-PAGE. Arrows indicate the spots that was ANF's and their details (#1-7) were described in table 2.

Fig. 2. Genotypic (G) variation of ANF's analyzed by PLS-DA model. Maryland (a: early, b: late), South Carolina (c: early, d: late), and South Dakota (e: early, f: late)

Fig. 3. Environmental (E) variation of ANF's from nine different soybean genotypes analyzed by PLS-DA. Maryland (a), South Carolina (b), and South Dakota (c)