Regulatory Toxicology and Pharmacology 58 (2010) S26–S29

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Natural variability in abundance of prevalent soybean proteins

Savithiry S. Natarajan *

USDA-ARS, Soybean Genomics and Improvement Laboratory, PSI, Beltsville, MD 20705, USA

ARTICLE INFO

Article history:
Available online 13 August 2010

Keywords:
Soybean
G. soja
G. max
2D-PAGE
MALDI-TOF-MS
LC–MS
Conglycinin
Glycinin

ABSTRACT

Soybean is an inexpensive source of protein for humans and animals. Genetic modifications (GMO) to soybean have become inevitable on two fronts, both quality and yield will need to improve to meet increasing global demand. To ensure the safety of the crop for consumers it is important to determine the natural variation in seed protein constituents as well as any unintended changes that may occur in the GMO as a result of genetic modification. Understanding the natural variation of seed proteins in wild and cultivated soybeans that have been used in conventional soybean breeding programs is critical for determining unintended protein expression in GMO soybeans. In recent years, proteomic technologies have been used as an effective analytical tool for examining modifications of protein profiles. We have standardized and applied these technologies to determine and quantify the spectrum of proteins present in soybean seed. We used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), and liquid chromatography mass spectrometry (LC–MS) for the separation, quantification, and identification of different classes of soybean seed proteins. We have observed significant variations in different classes of proteins, including storage, allergen and anti-nutritional protein profiles, between non-GMO cultivated and wild soybean varieties. This information is useful for scientists and regulatory agencies to determine whether the unintended expression of proteins found in transgenic soybean is within the range of natural variation.

1. Introduction

Soybean [*Glycine max* L. (Merr.)], is an important commercial crop grown as a high-protein feed supplement for its edible oil and for its other phytonutrients (Boulter et al., 1980). This has generated significant interest by major agricultural research institutes and biotech industries in enhancing the quality of nutrients in soybeans as well as in increasing resistance against various pests. This has resulted in the introduction of several new GMO cultivars in the US. As more transgenic cultivars become available, it has become important to identify and characterize the natural variation of proteins in non-genetically modified soybean genotypes as a baseline against which to compare the new cultivars.

At present, proteomics tools such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry, and liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) provide alternative methods to effectively examine protein profiles. These tools can be used to evaluate protein expression and potential protein modifications, as well as the relative abundance of proteins in non-GMO and GMO crops (Corpillo et al. 2004; Lehesranta et al. 2005). However, protein extraction remains a challenge in the analysis of proteins because of contaminants that could affect the performance of the 2D-PAGE, MALDI-TOF and LC–MS/MS analyses. In our laboratory, we carried out a comparison of four commonly used protein extraction methods described in the literature (Hurkman and Tanaka, 1986; Berkelman et al., 1998; Herman et al. 2003) for 2D-PAGE analysis, and subsequently selected an efficient extraction method for our analysis (Natarajan et al. 2005). This report summarizes some of the results obtained in our laboratory over the past few years. We have reported the variation of different classes of soybean seed protein profiles in cultivated and wild soybean genotypes. The wild genotypes are completely cross compatible with cultivated soybean by conventional plant breeding. Results from the wild genotypes can be compared with the amounts of storage, allergen and anti-nutritional proteins in GMO soybeans to determine if the levels of these proteins in the GMO soybeans are within the limits of natural variation.

2. Proteomic analysis of soybean storage proteins

The soybean storage protein families, β-conglycinin and glycinin, are classified based on their sedimentation coefficients.

* Address: Soybean Genomics and Improvement Laboratory, PSI-ARS-USDA, 10300, Baltimore Avenue, Beltsville, MD 20705, USA. Fax: 301-504-5728.
E-mail address: savi.natarajan@ars.usda.gov

0273-2300/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.yrtph.2010.08.005

β-conglycinins are encoded by two mRNA groups, the first mRNA group encodes α and α′ β-conglycinin subunits and the second mRNA group encodes the β-subunit of β-conglycinin (Schuler et al. 1982; Koshiyama, 1983). Initially, we used pH 3.0–10.0 in the first dimension to separate the subunits of storage proteins, however, this did not adequately separate glycinin storage protein subunits. Therefore, we currently used two narrow pH strips, 4.0–7.0 for acidic proteins (Fig. 2A and B) and pH 6.0–11.0 range for basic proteins (Fig. 3A and B). Triplicate samples were used for soybean seed protein extraction and 2D-PAGE analysis. Densitometry images were analyzed with Image Master 2D-Elite software (version 4.01) (GE Healthcare, Piscataway, NJ). Image analysis included the following parameters such as spot detection, background subtraction, spot measurement, and spot matching. A total of forty-seven different protein spots could be excised from 2D-PAGE gels using all three pH ranges and were characterized using mass spectrometry. Protein spots that could not be identified by MALDI-TOF-MS were further analyzed by LC–MS. All the proteins identified by MALDI-TOF-MS and LC–MS in the NCBI nonredundant database are listed in Table 1, which gives: an assigned protein spot number and its theoretical isoelectric point (pI) and molecular weight (Mr), protein identity, number of peptides matched, percent sequence coverage, MOWSE score, expect value, the NCBI database accession number of the best match, and a database that yielded concurrent identification.

Comparison of proteins from the wild and the cultivated soybean genotypes by 2D-PAGE using pH 3.0–10.0 IPG strips (Fig. 1A and B) highlight three subunits of β-conglycinin, the α, α′ and β-subunits (Natarajan et al. 2006, 2007a). The α subunit showed five protein spots (#1, 3, 4, 6 and 7) in the cultivated genotype (Fig. 1B) and seven spots (#1–7, Fig. 1A) were resolved in the wild genotype. The intensity of # 6 and #7 spots were higher in the wild genotype compared to the cultivated genotype. Some of the spots demonstrated heterogeneity in molecular weight and isoelectric point (pI), and these differences could be the result of post-translational modifications, proteolysis, and/or alternate splicing. The α′ subunit of β-conglycinin showed one spot (#8) in both wild and cultivated genotypes (Table 1), but the intensity of this spot was lower in the wild (Fig. 1A) compared with the cultivated genotype (Fig. 1B). The β-subunit of β-conglycinin resolved into four protein spots, (#9, 11–13) in the wild genotype and five protein spots (#9–14) in the cultivated genotype. Schuler et al. (1982) reported that β-conglycinin subunits are products of a multigene family, and the variation in the distribution of protein spots in our study could also be due to post-translational modifications (Koshiyama, 1983; Davies et al. 1985).

Glycinin is composed of five subunits, G1, G2, G3, G4 and G5, the precursors of which are encoded by five non-allelic genes, Gy1, Gy2, Gy3, Gy4 and Gy5, respectively (Nielsen et al., 1989). On pH 3.0–10.0 gels, it could be separated into acidic and basic polypeptides. However, many did not separate well. We now use narrow pH range IEF strips (4.0–7.0 and 6.0–11.0) to separate and identify the acidic and basic chains of glycinin subunits. The G1 subunit showed three basic polypeptides (spot #15, 16 and 17) in the pH 6.0–11.0 range that were similar in both wild and cultivated genotypes (Fig. 3A and B). The glycinin G2 showed eight spots (#18–25) of acidic and basic polypeptides in the wild genotype and seven spots in cultivated genotype. G3 glycinin acidic and basic polypeptides resolved into seven spots (#26–32) in the wild genotype, and the cultivated genotype showed an absence of two spots (#29 and 30), weak intensity of spot #28, and the presence of an additional spot#33 (Figs. 2 and 3 and Table 1). Two acidic polypeptides, which were identified as glycinin G4 (#34, 35), were present in the wild genotype and absent in the cultivated genotype (Fig. 1A and B). Glycinin G4 basic polypeptide separated into five spots (#36–40), was only present in the wild type seeds, and showed strong intensity (Fig. 1). The absence of G4 subunits in these particular cultivated genotypes could be due to the absence of the gene(s) encoding G4 subunits. Nielsen et al. (1989) reported the absence of G4 in soybean cv. Raiden, and suggested that heterogeneity of the G4 subunit may not have functional relevance. The glycinin G5 polypeptide (spot #41) that was clustered in the pH 3.0–10.0 gels (Fig. 1A and B) was clearly separated into three polypeptides (spot #42–44) in the narrow pH 4.0–7.0, in both wild and cultivated genotypes (Fig. 2). Three basic G5 polypeptides (spot #45–47) showed strong intensity in the cultivated genotype and weak intensity in the wild genotype. Fukazawa et al. (1985) reported the diversity of the glycinin G5 mRNA subunit between *G. max* and *G. soja*. Our study showed considerable variations of both acidic and basic glycinin polypeptides between wild and cultivated genotypes. The large amount of variation in seed protein composition between wild and cultivated genotypes may be due to a different complement of genes in the wild genotype that control expression of β-conglycinin and glycinin protein composition compared with the cultivated genotypes (Kwanyuen et al. 1997).

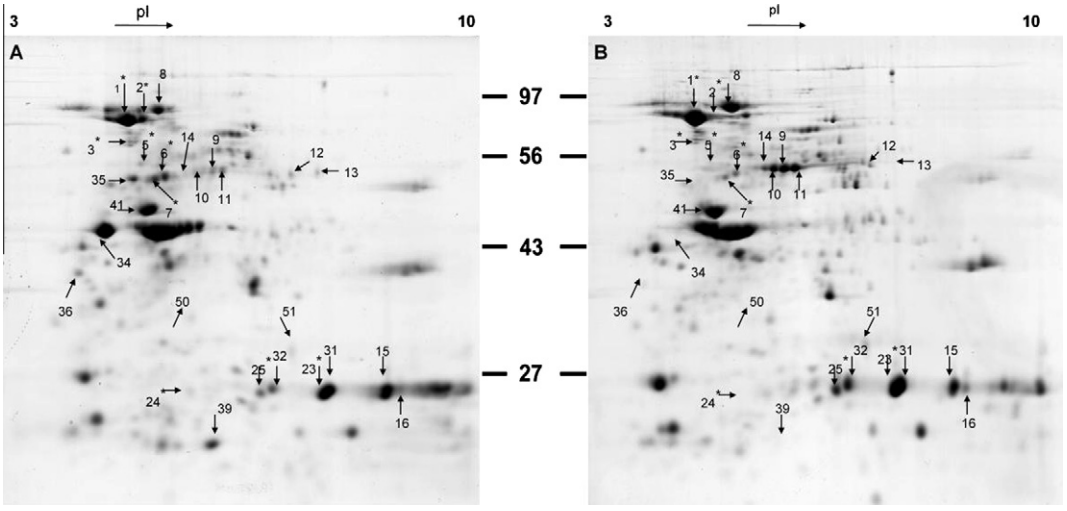

Fig. 1. Two-dimensional electrophoresis profile of soybean seed proteins of wild, *G. soja* (1A) and cultivated, *G. max* (1B). The first dimension IEF was performed using pH 3.0–10.0 linear IPG strips. Numbered arrows indicate the polypeptides referred to in the text and table.