Research Article

Characterization and differentiation of diverse transgenic and nontransgenic soybean varieties from CE protein profiles

Nowadays, soybeans are commercialized in a wide variety of colors and tones. Moreover, some pigmented seeds are being commercialized as soybeans while, on other occasions, these seeds are labeled as mung beans, azuki beans or soybean frijoles generating confusion on their identity. In this work, CE has been applied for the first time for the characterization and differentiation of different pigmented beans commercialized as soybeans. Other seeds commercialized as azuki, mung green soybeans or soybean frijoles were also analyzed. Borate buffer (at pH 8.5) containing 20% v/v ACN was used as the separation media and solution containing ACN/water (75:25 v/v) with 0.3% v/v acetic acid was used to solubilize the proteins from the samples. A 50 cm bare fused-silica capillary was employed for obtaining adequate separations in about 12 min. The CE protein pattern observed for yellow soybeans was different from that corresponding to green and red soybeans. The seeds commercialized as black soybean presented electropherograms identical or similar to those yielded by the yellow seeds with the exception of the sample labeled as black soybeans frijoles that presented a totally different pattern. In addition, CE protein profiles obtained for azuki and mung green soybeans were very similar to those corresponding to red soybeans and green soybeans, respectively. Finally, the CE method was also applied to differentiate transgenic and nontransgenic soybean varieties. Discriminant analysis, using several protein peak areas as variable, was used to successfully classify these samples.

Keywords:
CE / Pigmented soybean / Protein profiles / Transgenic soybean

DOI 10.1002/elps.200600799

1 Introduction

Soybean is commercialized in a wide variety of colors and tones: from yellow to black passing through green, brown or red [1–6]. The most commercial and abundant soybean is the yellow pigmented one. Although, in principle, the other colored seeds are also considered as soybean, it is not very clear whether they are different varieties of soybean or different legumes. For example, green-coat soybeans could not be real soybeans but a different legume originally from China called mung bean (Vigna radiata (L) R. Wilezek). Another example is red-coat soybeans which could be mixed up with the azuki bean (Vigna angularis (Willd.) Ohwi and Ohashi), a sweet flavor bean very much used in confectionery [7, 8]. Similarly, black soybeans are commercially available as black soybeans or as black soybean frijoles. Therefore, the use of the word ‘soybean’ for the designation of different legumes can result in misuse generating some consumer confusion. In fact, soybean is a very valuable bean due to its nutritional and functional properties and these properties can be different from those of other legumes. In addition, genetic engineering has also yielded a further number of soybean varieties generated by the incorporation of certain genes into the original soybean genome. The genetic modification of soybean crops resulted in transgenic soybean. For instance, the transgenic glyphosate-tolerant soybean (called Roundup Ready) is a genetically modified soybean resistant to glyphosate herbicide used to kill nearby weeds without killing the genetically modified crops, commercialized first by Monsanto in 1998. To date, although transgenic soybean is allowed for human consumption in the European Union (EU), the cultivation of the beans is not approved yet (http://www.monsanto.com). Development, growth and release of new genetically modified seeds and their labelling when used as foods are regulated by the EU [9].

Correspondence: Professor Mª Luisa Marina, Dept. of Analytical Chemistry, Faculty of Chemistry, University of Alcalá, E-28871 Alcalá de Henares, Madrid, Spain
E-mail: mluisa.marina@uah.es
Fax: +34-91-8854971

Abbreviations: IE-HPLC, ion-exchange chromatography; GMO, genetically modified organism
Soybean cultivar differentiation has traditionally been performed based on the pigmentation and morphological traits. More accurate methods were based on the study of DNA and proteins present in soybean. The identification of soybean cultivars based on the analysis of soybean proteins (mainly glycinin and conglycinin) has mainly been carried out by using gel electrophoresis [10–15] and chromatographic [16, 17] techniques. RP [16, 17] and size-exclusion [17] chromatographies have been applied for the characterization of soybean cultivars. The chromatographic profiles enabled the differentiation among very different cultivars but it was very difficult to differentiate among close cultivars. Moreover, the analysis times employed for the separation of soybean proteins ranged from 20 to 90 min. More recently, our research group has applied RP chromatography (RP-HPLC) [18] and ion-exchange chromatography (IE-HPLC) [19] for soybean cultivar characterization. For that purpose, ten beans with different coat pigmentation (yellow, green, red and black) and marketed as soybeans were analyzed by RP-HPLC and IE-HPLC. In both cases, analysis times needed for the separation of soybean proteins (3 min by RP-HPLC and 9 min by IE-HPLC) were significantly lower than those previously employed by the other authors [16, 17]. Moreover, all seeds presenting the same coat color showed the same proteins profile by RP-HPLC with the exception of black soybeans. The application of IE-HPLC for the same purpose did not enable the clear differentiation among soybean cultivars because similar chromatograms were observed for all seeds. Regarding the detection of Roundup Ready soybean, the most frequent procedure is to apply PCR methods where DNA fragments are amplified and then usually determined by standard gel electrophoresis [20]. Very recently, other procedures combining PCR and CGE to analyse DNA fragments [21, 22] or 2-DE to study the protein profiles have been used as good approaches to assess the safety of food and feed derived from genetically modified crops [23–25].

CE is a powerful separation technique useful for the analysis of agricultural products and foods [26]. Thus, the identification of cultivars, quality control and detection of food adulterations are important issues that have already been carried out by CE of proteins [27, 28]. The goal of this work was to investigate for the first time the potential of CE to characterize soybeans commercialized as yellow, red, green or black soybeans based on their different protein profiles. Moreover, a study is carried out to investigate the usefulness of CE to differentiate between some transgenic and nontransgenic soybeans.

2 Materials and methods

2.1 Chemicals and samples

All reagents employed for the preparation of the separation buffers were of analytical grade. Boric acid was from Fluka (Buchs, Switzerland), sodium hydroxide and glacial acetic acids were supplied from Panreac (Barcelona, Spain) and HPLC-grade ACN was from Merck (Darmstadt, Germany). All solutions were prepared with ultrapure water from a Milli-Q system (Millipore, Bedford, MA, USA).

Sixteen beans commercialized as soybeans (see Table 1) and with different coat pigmentation (six yellow soybeans (S1, S2, S3, S4, S5 and S6), five green soybeans (S7, S8, S9, S10 and S11), one mung green soybean (S12), two red soybeans (S13 and S14), two azuki (S15 and S16), one black soybean frijoles (S17) and five black soybeans (S18, S19), were used for the CE analysis. These soybeans were obtained from local farmers and the commercial information was supplied by the suppliers. The samples were treated with 0.1 N hydrochloric acid and ethanol and then dried at 40°C.

Table 1. Seed morphological traits of the studied beans

<table>
<thead>
<tr>
<th>Commercial name</th>
<th>Pigmentation</th>
<th>Approximate size (cm)</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>White soybean (S1)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Yellow soybean (S2)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Soybean (S3)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Soybean (S4)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Soybean (S5)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Soybean (S6)</td>
<td>Yellow coat/yellow inside</td>
<td>~0.7</td>
<td>Round</td>
</tr>
<tr>
<td>Green soybean (S7)</td>
<td>Green coat/yellow inside</td>
<td>~0.5</td>
<td>Oval</td>
</tr>
<tr>
<td>Soybean (S8)</td>
<td>Green coat/yellow inside</td>
<td>~0.5</td>
<td>Oval</td>
</tr>
<tr>
<td>Green soybean (S9)</td>
<td>Green coat/yellow inside</td>
<td>~0.5</td>
<td>Oval</td>
</tr>
<tr>
<td>Green soybean (S10)</td>
<td>Green coat/yellow inside</td>
<td>~0.5</td>
<td>Oval</td>
</tr>
<tr>
<td>Green soybean (S11)</td>
<td>Green coat/yellow inside</td>
<td>~0.5</td>
<td>Oval</td>
</tr>
<tr>
<td>Azuki (S12)</td>
<td>Red coat/yellow inside</td>
<td>0.5–0.7</td>
<td>Oval</td>
</tr>
<tr>
<td>Red soybean (S13)</td>
<td>Red coat/yellow inside</td>
<td>0.5–0.6</td>
<td>Oval</td>
</tr>
<tr>
<td>Azuki (S14)</td>
<td>Red coat/yellow inside</td>
<td>0.6–0.8</td>
<td>Oval</td>
</tr>
<tr>
<td>Red soybean (S15)</td>
<td>Red coat/yellow inside</td>
<td>0.6–0.8</td>
<td>Oval</td>
</tr>
<tr>
<td>Azuki (S16)</td>
<td>Red coat/yellow inside</td>
<td>0.6–0.7</td>
<td>Oval</td>
</tr>
<tr>
<td>Black soybean frijoles (S17)</td>
<td>Black coat/white inside</td>
<td>~1.5</td>
<td>Kidney</td>
</tr>
<tr>
<td>Black soybean (S18)</td>
<td>Black coat/yellow inside</td>
<td>~0.8</td>
<td>Round/oval</td>
</tr>
<tr>
<td>Black soybean (S19)</td>
<td>Black coat/green inside</td>
<td>~0.8</td>
<td>Round/oval</td>
</tr>
</tbody>
</table>

a) Ecological cultivar.
S10 and S11), two red soybeans (S13 and S15) and three black soybeans (S17, S18 and S19) were purchased from local markets in Madrid (Spain). Other seeds commercialized as azuki were from local markets of Madrid (Spain) (S12 and S14) and Wien (Austria) (S16). Seed morphological traits of the studied samples are summarized in Table 1. The nontransgenic soybean flour was acquired from El Granero (Madrid, Spain). In addition, two soybean flours prepared from transgenic soybean were kindly provided by INIA (Madrid, Spain) (http://www.inia.es/) and 34 different yellow soybeans were given by a Soybean Germplasm Bank (Centro de Recursos Fitogenéticos, CRF, Madrid, Spain).

2.2 Apparatus

All experiments were performed on an HP 3D CE system (Agilent Technologies, Waldbronn, Germany) equipped with an on-column DAD. Instrument control and data acquisition were performed by the HP 3D CE ChemStation software. Separations were performed on untreated fused-silica capillaries of 75 µm id and 375 µm od with a total length of 58.5 cm and an effective length of 50 cm from Composite Metal Services (Worcester, England, UK). The selected instrumental conditions were: capillary temperature 25°C, injection by pressure, 50 mbar for 5 s, applied voltage 20 kV and UV detection at 210, 254 and 280 nm with a bandwidth of 5 nm in all cases and response time of 0.1 s. A 744 pH Meter from Metrohm (Herisau, Switzerland) was used to adjust the pH of the separation buffers.

2.3 Procedure

Conditioning of capillary consisted of flushing with 1 M sodium hydroxide (1 bar) for 30 min and with water (1 bar) for 10 min before first use. Between sample injections, capillaries were conditioned with 0.1 M sodium hydroxide (1 bar) for 2 min, Milli-Q water (1 bar) for 2 min and, finally, with the separation buffer (1 bar) for 2 min.

A 100 mM borate buffer at pH 8.5 was prepared by dissolving ~1.5 g of boric acid in Milli-Q water, adjusting the pH at 8.5 with 1 M sodium hydroxide and adjusting the final volume of the solution at 250 mL. This buffer solution (80 mL) was mixed with 20 mL of ACN to obtain the separation buffer solution: 80 mM borate buffer with 20% v/v ACN (apparent pH 9.2). Buffer solutions were filtered prior to use through 0.45 µm pore size disposable nylon filters from Titan (Eatontown, NJ, USA).

All beans were ground with an automatic miller. Protein extracts were prepared by dissolving the appropriate amount of the ground sample in ACN/water (23:75 v/v) containing 0.3% v/v acetic acid, the solubilization solution that provided successful results for soybean proteins extraction in the previous work of the authors [29]. Different sample concentrations were used for different soybeans studied to obtain similar detection signals since different pigmented soybeans contain different protein percentages m/m on dry basis (yellow soybeans ~40–50%, and red, green and black soybeans ~20–30% [18]). Final concentrations of 60 mg/mL for yellow beans (~0.3 g of sample in ~5 g of solubilization solution), 80 mg/mL for green and red beans (~0.3 g of sample in ~3.7 g of solubilization solution) and 100 mg/mL for black beans (~0.3 g of sample in ~3 g of solubilization solution) were employed. Soybean flour solutions were 40 mg/mL for nontransgenic soybean (~0.3 g of sample in ~7.5 g of solubilization solution) and 60 mg/mL for transgenic soybeans (~0.3 g of sample in ~5 g of solubilization solution). Finally, these solutions were manually shaken and centrifuged (3362 x g for 10 min, 25°C) before their injection into the CE system.

2.4 Data treatment

Peak areas were integrated by setting the baseline from valley-to-valley. The area percentage for every peak was calculated as the average of two replicates (injected by duplicate) using Excel Microsoft XP. Box-and-whisker plots and discriminant analyses were performed using the computer program Statgraphics Plus for Windows 5.0. Graphs with different electropherograms were composed in Origin® version 7.0 software.

3 Results and discussion

3.1 Characterization of different pigmented soybeans

A CE method previously developed by our research group to achieve the separation of soybean and rice proteins [29] was applied in this work for the characterization of soybeans. Borate buffer (80 mM) at pH 8.5 containing 20% v/v ACN was employed as the separation media. The solubilization medium for the preparation of protein extracts consisted of ACN/water (75:25 v/v) containing 0.3% v/v acetic acid. Figures 1 and 2 show the electropherograms corresponding to different soybeans as examples of every pigmentation (yellow, green, red and black) at three different wavelengths: 210 nm (absorption of peptide bonds), 254 nm (absorption of phenylalanine residues) and 280 nm (absorption of tyrosine and tryptophan residues). The electropherogram corresponding to one azuki sample (S12) was also included in Fig. 2 together with that obtained for a red soybean sample (S13). For yellow, green and black soybeans, the two electropherograms overlapped in every figure (three in the case of black beans) corresponding to those soybeans with the same coatcolor presenting the most different patterns. In all electropherograms, the first peak corresponded to EOF while the last peak appeared due to the presence of acetic acid in the solubilization medium used for the preparation of all sample solutions.

Clear differences between the electrophoretic patterns obtained for yellow soybeans and those corresponding to red, green or black soybeans could be observed at any of the three
Figure 1. Electropherograms corresponding to the injection of protein extracts obtained from yellow soybeans (~60 mg/mL) and green soybeans (~80 mg/mL) using a solution of water/ACN (75:25 v/v) with 0.3% v/v acetic acid. Separation conditions: 80 mM borate buffer at pH 8.5 containing 20% v/v ACN; uncoated capillary with 75 μm id and 50 cm effective length (58.5 cm total length); injection by pressure, 50 mbar for 5 s; capillary temperature, 25°C; applied voltage, 20 kV; UV detection at 210, 254 and 280 nm.
**Figure 2.** Electropherograms corresponding to the injection of protein extracts obtained from red beans (red soybeans and azuki) (~80 mg/mL) and black soybeans (~100 mg/mL) using a solution of water/ACN (75:25 v/v) with 0.3% v/v acetic acid. Other experimental conditions as in Fig. 1.
wavelengths used. In fact, when UV-absorption detection at 210 nm was used, nine peaks were observed for yellow soybeans, ten peaks for green soybeans, seven peaks for red soybeans and azuki and nine peaks for black beans. As expected, when using UV detection at 254 and 280 nm, selectivity increased while sensitivity and the number of peaks decreased (see Figs. 1 and 2).

Despite 210 nm was the wavelength that enabled to obtain the higher number of peaks, 254 and 280 nm are the wavelengths corresponding to the maximum absorption of certain common amino acids present in proteins. Since the aim of this work was the characterization and differentiation of soybean varieties from CE protein profiles, 254 and 280 nm were selected as detection wavelengths for further work.

In spite of the fact that proteins are the main components in soybean (from 40 to 50% m/m on dry basis in yellow soybeans and from 20 to 30% m/m for the other studied soybeans [18]), there are other minor components present in soybean that can also absorb at these wavelengths such as isoflavones. Although soybean is the most significant dietary source of isoflavones [30, 31], their content in soybean is about 0.4% m/m of the protein content [32]. In addition, the extraction of isoflavones is usually performed with very high percentages of organic solvents like ACN or methanol and we use only a 25% v/v ACN in our extraction medium for the extraction of isoflavones in CE and CEC. Although isoflavones are present in soybean, the use of ACN as an extraction solvent is not very important because the peaks observed had different wavelengths used. In fact, when UV-absorption detection at 254 nm enabled the separation of proteins in three main peaks for yellow soybeans and black soybeans S18 and S19 (corresponding to peaks 1, 4 and 5 marked at 210 nm) although for black soybeans S18 an additional big peak at about 7.5 min appeared that did not appear for the sample S19. Moreover, four main peaks for green soybeans (corresponding to peaks 4, 5, 6g and 7g marked at 210 nm) and red soybeans (corresponding to peaks 2r, 3r, 5r and 6r marked at 210 nm) were observed. A different nomenclature to design electrophoretic peaks was adopted for green and red soybeans with respect to yellow and black soybeans because the peaks observed had different migration times from those obtained for yellow and black soybeans. Finally, when UV detection was performed at 280 nm, the main peaks observed for yellow soybean and black soybeans (samples S18 and S19) were peaks 1, 2 and 4 although for black soybean S19 an additional peak was observed as for 254 nm. The main peaks observed for green soybeans at 280 nm were peaks 5g, 6g and 7g and in the case of red soybeans the main peaks were 2r, 5r and 6r. The comparison of peak area ratios at 254 and 280 nm for the selected peaks obtained for the different beans with the same pigmentation (yellow, red and green), showed that the detector response ratio could be considered constant (RSD <13% for major peaks; peak 5r of red beans presenting the highest variability, RSD ~25%). In addition, from electropherograms of Figs. 1 and 2, it can be observed that migration times of the different samples varied. In fact, the major variation of migration times was from sample to sample (~1–4%), whereas migration times repeatability for the same sample was reasonable (~1%, n = 4). Regarding peak area repeatability, the injection of sample four times consecutively, yielded an RSD of about 5% while peak area reproducibility in the injection of two individually extracted samples was about 3%.

Figures 1 and 2 also show that in general, electrophoretic patterns were characteristic for every pigmentation. In fact, the main difference among patterns corresponding to soybeans with the same coat color was the peak size. Thus, yellow soybeans could be characterized by a big peak at about 6.3 min (peak 4), green soybeans patterns could be characterized by two big peaks appearing between 7 and 8 min (peaks 6g and 7g) and red soybeans could be characterized by the presence of two overlapped peaks at approximately 7 min (peaks 2r and 3r at 254 nm) from which only peak 2r is observed at 280 nm. It is interesting to remark that although red beans were marketed as red soybeans or azuki (see Table 1), the electrophoretic patterns of the five red beans analyzed were quite similar. However, as stated before, in the case of black soybeans, two of the three black soybeans studied (samples S18 and S19) presented similar patterns that, at the same time, resulted similar to the patterns obtained from yellow soybeans, while the other sample (S17) showed a different profile. This sample was commercialized as black soybean frijoles and presented a kidney shape while the other black soybeans presented a round/oval shape. Similar results were observed for these samples by RP-HPLC [18]. From these results, it is possible to state that yellow and black soybeans seemed to be different varieties of soybean while green and red soybeans seemed to be very different soybean varieties or even other legumes since no significant differences were found among protein profiles obtained for azuki and red soybean samples or for green soybean and mung green soybean samples.

Differences among peak area percentages obtained for every peak and for the different beans were analyzed using the box-and-whisker plot. Although the differentiation among the pigmented beans studied in this work could be observed at any of the three wavelengths used, as stated before, 254 and 280 nm were selected since the differences were more clearly displayed due to the simplicity of the electropherograms. Figure 3 shows the box-and-whisker plots obtained for yellow (six samples), green (five samples) and red (five samples) beans at 254 and 280 nm. It can be observed that very similar plots were obtained for yellow soybeans at the two wavelengths considered, the greatest dispersion in peak area percentage being obtained for peaks 1 and 4. Moreover, the highest variability in peak area percentage for green beans corresponded to peak 5 (at 254 nm) and peak 6 (at 280 nm) and for red beans to peaks 3, 5 and 6 (at 254 nm) and 5 and 6 (at 280 nm), peak 6 at 254 nm for red beans being the one accounting for the maximum variability in peak area percentage.
The method was modified in order to use it as a fast screening method [33]. In fact, the electrophoretic pattern was obtained in a 8.5 cm capillary (injection by the outlet position) instead of using the 50 cm capillary (injection by the inlet position) used previously. Although, as expected, with an 8.5 cm capillary the separations were worse than those obtained with the longest capillary (see Fig. 4); the electrophoretic patterns obtained enabled the clear and fast differentiation among the different pigmented beans, thus making possible a significant reduction in the analysis time. Using a capillary voltage of 20 kV enabled obtaining protein profiles in only 70 s (12 times shorter than the analysis time employed using the initial CE method).

3.2 Discrimination between transgenic and nontransgenic soybeans

The same analytical method used for the characterization of the different pigmented soybeans was also applied to discriminate between transgenic and nontransgenic yellow soybeans. Figure 5 shows the electrophoretic profiles obtained for a commercial soybean flour, a yellow soybean and a transgenic soybean flour at 254 and 280 nm. When comparing the electropherograms corresponding to the commercial soybean flour and the transgenic soybean flour with that obtained for soybeans (S2), a significant increase in the area of peak 1 with respect to that of peak 4 was observed.
Since this increase was characteristic of soybean flours (transgenic and nontransgenic), it could not be attributed to the genetic modification of the transgenic soybean. As a consequence, peaks 2, 4 and 5 were selected in order to study if it was possible to characterize transgenic material using chemometric tools. Discriminant analysis was applied to the area percentage obtained for the selected peaks in nine samples that were distributed in the following categories: transgenic samples (genetically modified organism (GMO)), nontransgenic samples (nonGMO) and mixtures of transgenic and nontransgenic samples (mixtures) (see Fig. 6A). A 100% of cases correctly classified were obtained using the above-mentioned three peaks at 254 and 280 nm. Moreover, a 100% of correct classification was also obtained when the leave-one-out cross validation procedure was used. At 280 nm, even with only two variables (peaks 2 and 4), it was possible to correctly classify the nine samples of soybeans analyzed. Moreover, thirty-four new samples analyzed by the CE method were used to prove the validity of the discriminant model (see Fig. 6B where a total of forty-three samples were considered). A 93% of the cases were correctly classified as nontransgenic samples. Only three samples were incorrectly classified as a mixture of transgenic and nontransgenic soybeans although they were nonGMO samples.

4 Concluding remarks

This is the first time that CE has successfully been applied for the characterization of different pigmented soybeans as well as for the differentiation of transgenic and nontransgenic soybeans based on their protein profiles.
Electrophoretic patterns observed for the different pigmented soybeans were obtained in 12 min and enabled the clear differentiation between yellow soybeans and the other coloured soybeans. Yellow soybeans presented an electrophoretic pattern characterized by a huge peak at \( \sim 6.5 \) min that was similar to the pattern obtained for two of the black soybeans. Samples marketed as green soybeans, red soybeans and black soybean frijoles showed electrophoretic profiles characteristic for each pigmentation and different from those obtained for yellow soybean. These differences could be observed at 210, 254 and 280 nm, although were more clearly displayed at the two latter wavelengths due to the simplicity of the electropherograms. However, protein profiles obtained for mung green soybean and azuki were very similar to those obtained for green and red soybeans, respectively.

The modification of the method by injecting samples by the outlet position enabled a significant reduction in the analysis time to 70 s making this method useful as a fast screening tool. This method would enable an initial classification of the samples that, afterwards, could be characterized with more detail by the initial CE method.

Area percentages of peaks 2, 4 and 5 and the use of discriminant analysis were also successfully applied to the differentiation between transgenic and nontransgenic soybeans.

The authors thank the Comunidad Autónoma de Madrid (Spain) (project S-0505/AGR/0312) and the Ministerio de Ciencia y Tecnología (Spain) (project AGL2005-05120-C02-01/...
ALI) for financial support. C. García-Ruiz also thanks the Ministerio de Ciencia y Tecnología for the Ramón y Cajal program (RYC-2003-001).

5 References