CE-TOF MS analysis of complex protein hydrolyzates from genetically modified soybeans. A tool for Foodomics.

Carolina Simó[1], *Elena Domínguez-Vega*[2], *Maria Luisa Marina*[2], *María Concepción García*[2], *Giovanni Dinelli*[3], *Alejandro Cifuentes*[1,*]

[1] Department of Food Analysis, Institute of Industrial Fermentations, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

[2] Department of Analytical Chemistry, Faculty of Chemistry, University of Alcalá, Ctra. Madrid-Barcelona, Alcalá de Henares, Madrid, Spain

[3] Department of Agroenvironmental Science and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy

Running title: CE-MS of GM soybean peptides

Corresponding author: Alejandro Cifuentes, **e-mail:** acifuentes@ifi.csic.es; **Fax#** 34-91-5644853; **Phone#** 34-91-5618806 (ext. 315)

List of abbreviations: BPE, base peak electropherogram; EIE, extracted ion electropherograms; GM, genetically modified; IAA, iodoacetamide; TPCK, N-tosyl-L-phenylalanine chloromethyl ketone, SPI, soybean protein isolate; TUC, thiourea-urea-CHAPS-Tris buffer.

Keywords: Capillary electrophoresis, mass spectrometry, shotgun proteomics, transgenic foods.

Abstract

A CE-TOF MS proteomic approach was applied for the analysis of hydrolyzates from complex soybean protein mixtures. After CE-TOF MS method development, the new approach provided the simultaneous analysis of more than 150 peptides from the soybean protein fraction soluble in ACN-water (80/20, v/v). The method is fast (about 30 min of analysis per sample) and is characterized by a relatively low running cost. The approach was used to study the substantial equivalence between a genetically modified (GM) variety of soybean compared to its traditional counterpart. No significant differences were found between the two studied soybeans based on the protein fraction studied. The capacity of the CE-TOF MS method to analyze complex mixtures of peptides in short times opens interesting possibilities in the growing Foodomics area.

1. Introduction

The consumption of soybean derived products has increased considerably in the last years due to their good nutritional and health promoting properties and their use as ingredient in the elaboration of a large number of foods [1]. Genetically modified (GM) soybean is one of the main commercialized GM crops together with maize, cotton and canola. Due to the complex composition of foods, "safety assessment" for GM-crop-derived foods is not a straightforward task as already discussed by several authors in the literature [2]. In order to overcome this difficulty, the term "substantial equivalence" has been proposed to achieve a better understanding on the chemical composition of GM crops [3]. Substantial equivalence studies should enable the identification of potential differences between the GM crop-derived food and the traditional counterpart with a history of safe use during decades [4]. Targeted analyses have been mainly used to corroborate this substantial equivalence, demonstrating their usefulness for the study of the primary or intended effect of the genetic modification. However, targeted analyses fail to identify any other unintended effect generating important doubts about their convenience to characterize new genetically modified organism (GMOs). In our opinion, profiling technologies such as genomics, transcriptomics, proteomics, and metabolomics, should be the method of choice to exhaustively investigate the physiology of GM plants [5-9]. Moreover, recently the European Food Safety Agency (EFSA) has recommended the development of profiling analysis to extend the breadth of comparative analyses among GMOs and their non-transgenic counterparts [10].

Proteins are of special interest since they can e.g., have antinutrient properties, act as enzymes, have allergenic properties, be involved in the synthesis of toxins, etc.

Improved methods for proteins profiling may be helpful to discover unexpected changes in GM plants corroborating the substantial equivalence. In this regard, soybean is a high-quality source of vegetable proteins with low fat content and demonstrated health-promoting properties [11]. Seed protein concentration of commercial soybean cultivars is approximately 40% w/w (calculated on a dry weight basis). The two main storage soybean proteins glycinin (11S globulin) and β-conglycinin (7S globulin) constitute about 80-90% w/w of the total soybean proteome. The relative levels of these proteins which can vary among genotypes, maturity, etc., have a significant impact in final soy-based food nutritional and functional properties [12,13]. Protein profiling has also been demonstrated to be useful to carry out processing control studies and detection of soybean food adulteration [14].

In a series of works, Fernandez Ocaña *et al*. [15,16] demonstrated the potential of several target LC-MS approaches for the detection and characterization of the transgenic protein CP4 EPSPS in GM soybean and maize. Although the detection of newly expressed protein is important for the investigation of the intended effect that results from the genetic modification, protein profiling will increase the chances of detecting unintended effects. In addition to SDS-PAGE [17], and immunological methods [18], 2-DE combined with MS is still the most widely approach to compare entire plant proteomes (or subproteomes) to identify the differently expressed proteins [17,19]. However, several advanced methodologies such us LC [20], capillary electrophoresis [21] and multidimensional LC [22] have also been used for the profiling analysis of soybean proteins. Results have demonstrated the complexity of intact soybean protein separation. Thus, perfusion ion-exchange chromatography with UV detection was used for the separation of proteins from commercial soybeans, allowing

the characterization of these different wild soybean cultivars [20]. A perfusion reversed-phase HPLC-MS method was also developed enabling the analysis of intact soybean proteins in different soybean cultivars to study the similarities and differences among soybeans with different pigmentation [23]. CE with UV detection was also used for the characterization of different pigmented soybeans as well as for the differentiation of transgenic and non-transgenic soybeans based on their protein profiles [7]. Characterization of industrial soybean protein isolates has also been carried out without any previous separation step by direct analysis by MALDI-TOF MS [24]. In this case the composition of the different subunits of the most abundant proteins glycinin and β-conglycinin was used for a rapid fingerprinting of different industrial soybean protein isolates and a subsequent use of this methodology for quality control.

In general, the major difficulty in the analysis of proteins comes from their widely different physico-chemical properties (size, shape, charge and hydrophobicity). Furthermore, the large number of proteins and the huge differences in abundance in real samples make especially difficult their analysis. These inherent differences are usually exploited in fractionation procedures in order to decrease sample complexity and thus facilitating their subsequent separation by chromatographic and electrophoretic techniques. However, efficient separation of complex mixtures of proteins by automatic techniques such as LC and/or CE within a single analysis is still very difficult, and identification of large proteins by MS is challenging. In this sense, peptides are in general better separated by LC and/or CE techniques and also more easily analyzed by MS than proteins. For these reasons, an alternative approach to protein profiling is the so-called "shotgun proteomics" [25]. In this approach, a complex protein mixture is first digested with a suitable protease to generate peptides. For the analysis of such a

complex mixture of peptides, a previous separation step in two or more chromatographic and/or electrophoretic dimensions before MS analysis is required since the resolution of only a mass spectrometer is usually not sufficient. Current nLC–MS/MS methodology with instruments capable of delivering nanoliter/min flow rates, has been successfully applied to proteomics research and has provided dramatic improvement in protein identification. Capillary electrophoresis-mass spectrometry coupling is also becoming an important tool in shotgun proteomics [26] as a result of the today's increasing need of new high-throughput methodologies. There is an increasing number of works using CE-MS for the analysis of peptides for a multiple type of applications (clinical diagnosis, analysis of pharmaceuticals and related substances, environmental applications, food analysis, etc.) [27-30]. CE is a rapid and automated analytical technique that provides very high separation efficiencies, being complementary to LC since they are based on different separation mechanisms. Currently, there is an increasing number of proteomic applications using multidimensional separation systems based on hyphenation of CE, CEC, LC, and MS techniques to separate complex mixtures of peptides. [31]. On the other hand, the on-line coupling between CE and MS has already demonstrated a huge potential in many applications fields [28-30,32,33].

In this work, the use of CE-TOF MS is investigated as a complementary non-targeted tool to the existing analytical techniques to improve the probability of detecting unintended effects caused by a genetic modification. More precisely, a new CE-TOF MS method is developed for soybean shotgun proteomics giving additional evidences on the safety assessment of this GM crop compared with its traditional counterpart.

2. Materials and methods

2.1. Chemicals and samples

All reagents employed in the preparation of CE buffers and sheath liquids were of MS grade. Isopropanol, ACN and water were supplied by Riedel-de Haën (Seelze, Germany) and methanol was from Sigma (Steinheim, Germany). Ammonium hydroxide and ammonium hydrogen carbonate were purchased from Fluka (Deisenhofen, Germany), formic and acetic acid were from Riedel-de Haën. The buffers were stored at 4º C and warmed at room temperature before being used. All chemicas for SDS-PAGE analysis (Tris, HCl, SDS, glycine) were from Sigma. Gels for SDS-PAGE analysis were prepared with 40% v/v acrylamide/bis-acrylamide solution with a 37.5:1 w/w mixing ratio from Sigma. Gel staining solution was prepared with Coomassie from Sigma. 4-Dithiothreitol (DTT), iodoacetamide (IAA), bovine serum albumin (BSA) and protein standard mixture used as Mr marker in SDS-PAGE analysis, were from Bio-Rad (Hercules, CA, USA). Laemmli buffer (4% w/v SDS, 20% v/v glycerol, 10% v/v 2-mercaptoethanol, 0.004% w/v bromophenol blue, and 0.125 M Tris HCl, pH approx. 6.8) and trypsin TPCK (N-Tosyl-L-phenylalanine chloromethyl ketone) treated from bovine pancreas, were purchased from Sigma. All solutions were prepared with ultrapure water from a Milli-Q system from Millipore (Bedford, MA, USA).

The soybean protein isolate (SPI) with 89.1% w/w of soybean protein (determined by Kjeldahl method) used during the method development was from ICN (Aurora, OH, USA). Soybeans SB10 (transgenic) and SB11 (isogenic non-transgenic) used for the comparative peptide study were grown under the same conditions in a growth chamber.

Transgenic and non-transgenic nature of these soybean samples was confirmed based on their DNA using an analytical procedure developed in our laboratory and described elsewhere [34].

2.2. Protein extraction from soybean seeds

All beans were ground with a domestic miller prior to extraction. Protein extracts were prepared by directly dissolving 150 mg of ground sample in 1.5 mL of two different extraction buffers: the first buffer (TUC) contains 7 M urea, 2 M thiourea, 3% w/v CHAPS and 40 mM Tris base; while the second buffer contains ACN/water (80:20, v/v) as used by Garcia *et al*. [35]. After 10 min vortexing and 3 min sonication, extracts were centrifuged (9000 *g*) for 5 min. The supernatant fraction was collected and stored at -80º C until used. The total protein content was determined by the Bradford method using a commercial dye reagent from Bio-Rad (Hercules, CA) and using BSA as standard. All protein extracts were diluted to a concentration of 3 mg/mL before SDS-PAGE analysis (*vide infra*) or enzymatic digestion.

2.3. SDS-PAGE analysis

15 µL of the diluted protein extract was mixed with 15 µL of Laemmli buffer and heated in boiling water for 5 min. The mixture was then loaded in a 1.5 mm thick polyacrylamide gel. The SDS-PAGE gel was composed by a stacking gel with a large pore polyacrylamide gel (T=4%, C=2.6%) in 125 mM Tris-HCl, pH 6.8, 0.1% w/v SDS, over the resolving gel with a large pore polyacrylamide gel (T=14%, C=2.6%) in 375 mM Tris-HCl, pH 8.8, 0.1% w/v SDS buffer. The cathodic and anodic compartments

were filled with Tris-glycine buffer, pH 8.3, containing 0.1% w/v SDS. The electrophoretic run was performed by setting a voltage of 100 V until the dye front reached the bottom of the gel. Gel was stained in colloidal Coomassie brilliant blue and destained in 7% v/v acetic acid solution.

2.4. Enzymatic hydrolysis

Two protocols of enzymatic digestion were studied in this work. To carry out the first protocol, 50 μL of the diluted protein extract was boiled for 10 min. Then 50 μL of a freshly prepared bovine pancreas trypsin solution (in 50 mM NH_4HCO_3 at pH 8.0) was added at different enzyme-substrate ratios (1:50, 1:25 and 1:5, w/w), and the solution was incubated overnight at 37 ºC under continuous shaking (600 rpm) in a thermomixer from Eppendorf (Hamburg, Germany). The reaction was stopped by heating at 90 ºC for 5 min. Afterwards the suspension was centrifuged at 14000 *g* for 5 min, and the supernatant fraction was collected and stored at -20º C prior to analysis by CE-TOF MS. In the second protocol, the digestion was performed under reductive alkylation conditions to reduce disulfide bonds mixing 100 μL of the diluted protein extract with 5 μL of 200 mM DTT (in 100 mM NH_4HCO_3 at pH 8.0) and incubation at 90º C for 5 min. After allowing the sample to cool at room temperature, alkylation of free thiol groups with 4 μL of 1 M IAA (in 100 mM NH_4HCO_3 at pH 8.0) for 45 min at room temperature was carried out. To neutralize the remaining IAA, 5 μL of 800 mM DTT (in 100 mM NH_4HCO_3, pH 8.0) were added, and the solution was maintained at room temperature for 45 min. Finally, 50 μL of freshly prepared bovine pancreas trypsin solution (in 50 mM NH_4HCO_3 at pH 8.0) were added at different enzyme-substrate ratios (1:50 and 1:25, w/w), and the solution was incubated at 37º C overnight under

continuous shaking conditions in a thermomixer. The reaction was stopped heating at 90º C for 5 min, and the suspension was then centrifuged at 14000 *g* for 5 min. The supernatant fraction was collected and stored at -20º C prior to analysis by CE-TOF MS. Blank was performed under the same conditions by using the protein extraction solution plus digestion but without any protein.

2.5. CE-TOF MS analysis

Capillary electrophoretic analyses were carried out in a P/ACE 5010 CE apparatus from Beckman Instruments (Fullerton, CA, USA). The instrument was controlled by a PC running the System Gold software from Beckman. Uncoated fused-silica capillaries (50 μm id, 90 cm total length) from Composite Metal Services (Worcester, England) were coupled to MS through an orthogonal electrospray ionization (ESI) interface model G1607A from Agilent Technologies (Palo Alto, CA, USA) to the MS. ESI allows measuring the masses of large molecules by producing multiply charged ions, thereby decreasing the m/z till detectable ranges. A time-of-flight micrOTOF MS instrument from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a personal computer running the micrOTOF control software from Bruker Daltonics.

Before first use, the separation capillary was conditioned by rinsing with 1 M NaOH for 10 min, followed by 20 min with Milli-Q water and 5 min with the separation buffer. After each run, the capillary was conditioned with separation buffer for 4 min. Injections were made at the anodic end using N_2 presure of 0.5 psi (34.5 mbar) for 20 s. The electrophoretic separation was achieved using 25 kV as running voltage at a

constant temperature of 25 ºC. Electrical contact at the electrospray needle tip was established via a sheath liquid based on isopropanol-water (50:50, v/v) and delivered at a flow rate of 3 µL/min by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). The mass spectrometer operated with the ESI source in the positive ion mode. The nebulizer and drying gas conditions were 0.4 bar N_2 and 4 L/min N_2, respectively, and maintaining the ESI chamber at 200º C. The micrOTOF was operated to acquire spectra in the range of 50–3000 m/z every 90 µs. The accurate mass data of the molecular ions were processed by DataAnalysis 3.3 software (Bruker Daltonics). External calibration of the TOF MS instrument was performed by introducing a 2% v/v Tuning Mix solution from Agilent through the separation capillary (by applying a pressure of 20 psi (1.38 bar)) towards the TOF MS instrument. Masses for the calibration of the TOF MS instrument were: 322.0481, 622.0209, 922.0098, 1321.9842, 1521.9715 and 2121.9332 m/z. Unless stated in text, all CE-TOF MS injections were made by duplicate and the abundance cutoff for mass spectra deconvolution was initially set at 10%.

3. Results and discussion

3.1. Protein extraction

Considering the great diversity and heterogeneity of proteins, simultaneous solubilization of all proteins from a certain tissue is still a challenge. In this work, two different protein solubilization solutions were used to study soybean proteins. First solution (TUC) contains chaotropes (urea and thiourea) and the detergent CHAPS, while the second solution is composed by ACN-water (80:20, v/v), as previously used by García *et al*. [35]. Although, total protein concentration of the extract solubilized

with TUC solution (25.2 mg/mL) was considerably higher that that obtained with the hydroorganic solution (3.2 mg/mL), the protein profiles were practically the same for both extraction process. Thus, as can be see in Figure 1, the predominant presence of the storage proteins β-conglycinin and glycinin was detected in both extracts.

The good quality of the protein extract obtained with the ACN/water mixture and its simplicity of preparation took us to select this procedure for further experiments. Moreover, an additional advantage of using ACN/water is to prevent possible interferences of the high concentration of the chaotrophic agents of the TUC buffer on the subsequent hydrolysis with trypsin. In a previous paper [35], our group demonstrated that ACN/water extracts can be used together with different LC-UV methods to obtain profiles useful to classify different soybean samples, including transgenic soybeans. However, only UV absorption at 280 nm was applied in that work [35] and, therefore, only relatively limited information about the extract composition could be obtained. Besides, these extracts were shown to be relatively unstable probably due to the unstable nature of the numerous proteins in the extract and the presence and action of proteolytic enzymes.

In order to obtain additional evidences on the safety assessment of this GM crop compared with its traditional counterpart, in this work, soybean samples are compared through the CE-MS analysis of the more stable peptidic fraction obtained from the enzymatic hydrolysis of the selected protein extracts. To do that, a new shotgun proteomics approach has to be developed in order to establish that comparison.

3.2. Optimization of tryptic digestion of soybean proteins

Trypsin, a pancreatic serine endoprotease, was selected in this work for soybean protein hydrolysis due to its highly specific cleavage at the C-terminal side of lysine and arginine, being the cleavage low-dependent on duration and conditions of digestion. The commercial enzyme used in this work has been exhaustively processed by reductive methylation to minimize autolysis and chymotryptic activity quenched by TPCK treatment. To carry out a suitable hydrolysis of the soybean proteins in the extract, a study of the best reaction conditions was performed taking into account the presence of other possible trypsin inhibitors in the sample. For this reason, two protocols of enzymatic digestion based on denaturalization of possible inhibitors of trypsin were investigated (see section 2.4). First, the use of high temperatures prior to digestion to inactivate trypsin inhibitors and increase protein digestibility, was studied [36]. Figure 2A shows the base peak electropherograms (BPE) obtained by CE-TOF MS of the different soybean protein hydrolysates (continuous line) and their corresponding blanks (dotted line) at three different trypsin-protein ratios (namely, 1:5, 1:25 and 1:50 w/w). It can be observed that heat treatment alone is not sufficient to completely inactivate trypsin inhibitors in the soybean protein extracts, and it is necessary to increase the trypsin-protein ratio to 1:5 w/w to observe a peptidic profile (upper electropherograms in Figure 2A). It was also found that using an enzyme:substrate ratio of 1:5 w/w the autodigestion of the enzyme increased and the peptides from trypsin (dotted line) interfered significantly with the analysis of the peptide map from the SPI extract (continuous line), as can be seen in the two upper electropherograms in Figure 2A. For this reason, the use of hydrolysis with a reductive alkylation protocol was investigated in order to inactivate trypsin inhibitors and increase protein digestibility. Although reduction and alkylation steps are generally used to favor the accessibility of the

substrate in the enzymatic proteolysis, this method can also be used to inactivate trypsin inhibitors because the reduction of disulfide bonds alters protein structure reducing inhibitor activity [37]. The digestion under reductive alkylation conditions was performed as described under section 2.4, using two different enzyme-substrate ratios (1:50 and 1:25, w/w). In this case, a 1:5 trypsin-protein ratio was not investigated in order to skip the interferences coming from the auto-hydrolytic fragments from the enzyme at high concentration as shown in Figure 2A (upper dotted line). After digestion using the reductive alkylation protocol, the soybean protein hydrolysates obtained were analyzed by CE-TOF MS and the results are shown in Figure 2B. As it can be observed in Figure 2B, the use of reductive alkylation protocol increased the protein digestibility allowing to obtain a good peptide map without interferences of auto-hydrolytic fragments from trypsin (dotted line). On the other hand, as can be seen in Figure 2B no significant differences between both trypsin-protein ratios (1:50 and 1:25, w/w) were found, therefore, a ratio of trypsin-protein 1:25 w/w was selected. In order to demonstrate the repeatability of the digestion procedure developed in this work, five different hydrolysis of the SPI extract were carried out under these conditions (i.e., using reduction and alkylation conditions and a 1:25 w/w trypsin-protein ratio). After CE-TOF MS analysis of the enzymatically digested protein extracts, the same set of peptides were found in all of them, so that, we apply this digestion protocol to carry out the enzymatic digestion of the protein extracts from the soybeans studied.

3.3. Optimization of CE-TOF MS conditions

The coupling between CE and MS requires the use of volatile BGEs compatible with electrospray ionization. In this work, volatile BGEs were studied together with low pH

buffers for CE-TOF MS analysis of the peptides from soybean samples. At extreme low pH values, the silanol groups of the capillary wall are protonated, minimizing peptide adsorption onto the inner capillary wall. Moreover, practically all peptides obtained from tryptic digestion should have at least two amine residues. Thus, at very low pHs the peptides should be positively charged (since the acidic groups would remain non-ionized) and these biomolecules will migrate toward the cathode. However, the main drawback of the use of extreme low pH BGEs is the decreasing resolution among some peptides by diminishing their mobility differences. BGEs containing acetic acid or formic acid at different concentrations (from 0.5 M to 1.5 M) were studied in order to obtain best separations in the shortest analysis time. Better results in terms of signal/noise ratio were obtained using formic acid based electrolytes instead of acetic acid and, therefore, formic acid was selected for further experiments. Figure 3 shows the CE-TOF MS electropherograms from the tryptic digest of the SPI extract obtained using BGEs containing different concentrations of formic acid. For a better understanding of the effect of formic acid concentration, some characteristic peaks are marked with letters from A to H in Figure 3. When increasing the concentration of formic acid it could be observed a gradual decreasing in the migration times. On the other hand, it was also observed a progressive decrease of the number of peaks. Namely, 240, 215, 190 and 150 peptides were detected by CE-TOF MS using 0.25 (pH 2.16), 0.5 (pH 2.02), 1 (pH 1.82) and 1.5 M (pH 1.69) formic acid BGEs, respectively. In summary, BGEs with lower concentration of formic acid allowed the detection of a higher number of peaks at the expense of a much higher analysis time. This behaviour is explained considering that higher concentrations of formic acid bring about lower pH values, increasing the positive charge of peptides and as a result, increasing their cathodal electrophoretic mobility, bringing about shorter analysis times and higher comigration during

separation. Due to the complexity of the sample, in *shotgun* proteome analysis by CE-MS not all comigrating peptides at a given analysis time will have optimum ionization conditions since the presence of strongly ionizable peptides will suppress the signal of the less ionizable ones. Moreover, complexity of the MS spectra is logically higher when comigration occurs. For these reasons, the higher the extent of comigration the lower the number of detected peptides. In this work, a BGE composed by 0.5 M formic acid was finally selected as a good compromise between number of detected peptides and analysis speed (see Figure 3C).

It is well-known that sheath liquid has also a significant effect on robustness and sensitivity of CE-MS analysis when using a sheath-flow ESI interface. Thus, the type of organic solvent present in the sheath liquid was first studied in this work. Different mixtures of ACN, methanol, and isopropanol with water were tested. It was observed that a 50:50 (v/v) isopropanol-water mixture provided the best response in terms of sensitivity and signal stability. The addition of formic or acetic acid to this solution was not studied due to the high quantity of formic acid already present in the separation buffer. In our experience, this quantity is more than enough to ensure peptides ionization in the ESI interface. In addition, the spray was shown to be stable using a sheath liquid flow rate between 2 and 6 µL/min, reaching the best S/N value at 3 µL/min. Finally, the influence of the nebulizer pressure (0.2-0.7 bar) was studied. The best results in terms of MS sensitivity were obtained using a nebulizer pressure of 0.4 bar. Other ESI parameters selected were dry gas flow of 4 L/min and dry gas temperature of 200 ºC.

Figure 4 shows the CE-TOF MS electropherogram obtained for the tryptic digest of a

SPI extract under the selected analytical conditions. Peptide mixture was analyzed in 32 min with good efficiencies (e.g., 238.000 plates/m for peak F). Tryptic peptides are usually assumed to be doubly charged as a result of the two basic sites. It was observed that the ESI interface produced in a considerable number of cases, triply and quadrupled charged peptides. Some examples are shown in Figure 5, in which the mass spectra and the resulted neutral mass spectra after deconvolution are present for some representative peaks of Figure 4, namely, A, C and G. In some cases, up to 5 species could be detected in the same peak (see, for example, Figure 5, panel C), showing that still some peptides comigrate, although in this case the problem could be tackled through the MS capabilities.

In order to show the complexity of the hydrolyzed SPI, the extracted ion electropherograms (EIE) of a number of detected peptides are shown in Figure 6. Due to the rapid spectral acquisition of TOF MS with enhanced mass resolution (exceeding 10000 FWHM), this mass analyzer is well-suited for on-line coupling with fast separation techniques like CE to analyze complex samples such us protein hydrolyzates from foods. These results show that CE-TOF MS can be very useful for many other applications in the new field of Foodomics. In this sense, it can be briefly mentioned that several attempts were carried out to analyze the same protein hydrolyzate employing a CE-MS with a standard ion trap MS analyzer (CE-IT MS). The shape of the electropherogram obtained by CE-IT MS was analogous to that obtained with the CE-TOF MS. However, when comigration of several peptides occurred, no useful information could be obtained from the IT MS spectra, mainly due to the lower resolution of the IT MS. However, the possibility of carrying out MS^n experiments with the IT MS analyzer has to be considered has an important advantage over the single

TOF MS analyzer for peptides sequencing.

Repeatability of the CE-TOF MS method was next studied for 6 of the entire set of peaks (named as A to H in Figure 4) within the same day (five consecutive injections) and three different days (n=15), obtaining %RSD values for analysis times lower than 1.6 % within the same day and lower than 2.5% in the inter-day study, corroborating the usefulness of this approach.

3.4. CE-TOF MS analysis of digested soybean protein extracts

The hydrolyzed protein extracts from conventional and GM soybean were analyzed by CE-TOF MS and the results are shown in Figure 7. CE-TOF MS electropherograms of Figure 7 show peptide profiles different to that obtained from the commercial extract (i.e., soybean protein isolate, SPI) shown in Figure 4. Namely, a slightly lower number of peptides seem to be detected in the extracts from conventional and GM soybeans of Figure 7 compared to that from the commercial SPI of Figure 4. This can probably be due to the higher complexity of the "real" soybean extracts compared to the more purified SPI sample. Thus, depending on the methodology used for the protein isolation (isoelectric precipitation, centrifugation, precipitation, etc.) [38], the protein content of the SPI could be modified compared with the protein content of soybean not subjected to any subsequent processing. Moreover, other soybean metabolites that have low molecular masses are expected to be present in the extracts from conventional and GM soybeans [39]. The presence of these species in the ESI chamber could generate certain suppression of peptides ionization leading to signal suppression.

To carry out an accurate comparison between conventional and GM soybean peptide

profiles of Figure 7, five consecutive CE-TOF MS analysis of the hydrolyzed protein extract from conventional soybean were carried out. The overlapped peptide signals in the obtained mass spectra were deconvoluted in order to determine the uncharged mass of the biomolecules. The number of mass spectra to be processed in a single electropherogram was huge, therefore, the only alternative was to carry out an automated interpretation. To do this, the abundance cutoff for automatic detection of peptides was initially set at 5%. This parameter indicates the minimum percentage abundance of an ion, relative to the most intense ion in the spectrum that will be considered for the deconvolution, determining which m/z signals from the mass spectra will be used to find compounds. Although using an abundance cutoff of 5% a high number of peptides could be detected (more than 200), some of the peptides were not reproducibly found in all the replicates. Therefore, higher abundance cutoff percentage was specified in order to eliminate unstable signals from non-abundant peptides, at the expenses of obtaining a lower number of peptides and consequently less information. Thus, using a cutoff of 15% a reproducible set of 151 peptides was detected in all the five samples. Next, the mass list produced using the cutoff value of 15% was used for the comparison between conventional and GM soybean.

As can be seen in Figure 7, comparing the two samples (conventional and GM soybean) it can be deduced that both CE-TOF MS electrophoretic profiles seem to be very similar. For a deeper comparison, masses of every detected peptide were calculated from the obtained mass spectra as indicated above. Peptides with molecular masses from 452.5 to 5178.8 Da were found in both samples. No differences between the 151 peptides detected were observed using the 15% cutoff and automatic deconvolution of the detected ions. These results seem to corroborate the equivalence between GM

soybean and its conventional counterpart. It should also be mentioned that other less abundant peptides were detected showing differences between the conventional and transgenic soybean. However, as mentioned before, these peptides were not included in the comparison as they did not fit within the repeatability constraint established for this study (i.e., all detected peptide must be detected in five consecutive injections). These results indicate that more work is required to further establish the equivalence between GM and conventional foods.

4. Concluding remarks and future outlooks

In this work, a shotgun proteomic approach based on the use of CE-TOF MS for the characterization of complex soybean protein hydrolyzates was developed. CE-TOF MS was investigated as a complementary non-targeted tool to the existing technologies to improve the probability of detecting unintended effects caused by a genetic modification. The power of CE-TOF MS methodology allowed the simultaneous analysis of more than 150 peptides from an ACN-water soluble soybean protein fraction in an automated and fast way with low operation cost. Results showed that in the soybean proteome obtained with an ACN-water mixture no qualitative changes were found between the conventional and the transgenic soybean.

This is the first time that CE-MS has been applied to the characterization of soybean through the analysis of digested protein fractions. Further profiling studies based on CE-MS will be carried out to better understand the consequences of genetic manipulation and to elucidate potential variations in seed composition. In this regard, the use of CE-TOF MS opens new possibilities in the new Foodomics field. For instance, it can also

be an effective tool for Nutrigenomics studies (e.g., biomarker detection in biological fluids after the ingestion of a determined bioactive ingredient or functional food), study of new recombinant enzymes to improve functional properties of the generated peptides (antihypertensive, immunomodulatory, antioxidant, etc.).

Acklowledgements

This work was supported by AGL2008-05108-C03-01 and CTQ2006-03849/BQU projects (Ministerio de Educacion y Ciencia, Spain) and CSD2007-00063 FUN-C-FOOD (Programa CONSOLIDER, Ministerio de Educacion y Ciencia, Spain) projects. E. D-V. thanks the University of Alcala for her pre-doctoral grant.

References

[1] Isanga, J., Zhang, G.N., *Food Rev. Int.* 2008, *24*, 252-276.

[2] Kuiper, H.A., Kleter, G.A., *Trends Food Sci. Technol*. 2003, *14*, 277-293.

[3] Organization of Economic Cooperation & Development (OECD). 1993. Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles, Organisation for Economic Co-operation and Development, Paris, (http://www.oecd.org/pdf/M00034000/M00034525.pdf).

[4] Kok, E.J., Kuiper, H.A., *Trends Biotechnol*. 2003, *21*, 439-444.

[5] Garcia-Cañas, V., Cifuentes, A., Gonzalez, R., *Crit. Rev. Food Sci*. 2004, *44*, 425-436.

[6] Garcia-Cañas, V., Gonzalez, R., Cifuentes, A., *TRAC-Trend Anal.Chem*. 2004, *23*, 637-643.

[7] García-Ruiz, C., García, M.C., Cifuentes, A., Marina. M.L., *Electrophoresis* 2007, *28*, 2314-2323.

[8] Garcia-Villalba, R., Leon, C., Dinelli, G., Segura, A., Fernández, A., Garcia-Cañas, V., Cifuentes, A., *J. Chromatogr. A* 2008, *1195*, 164-173.

[9] Levandi, T., Leon, C., Kaljurand, M., Garcia-Cañas, V., Cifuentes, A. *Anal. Chem*. 2008, *80*, 6329-6335.

[10] European Food Safety Agency. 2006. Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants and derived food and feed. EFSA Communications Department, Parma, Italy.

[11] Mateos-Aparicio, I., Redondo Cuenca, A., Villanueva-Suarez, M.J., Zapata-Revilla, M.A., *Nutr Hosp*. 2008, *23*, 305-312.

[12] Riblett, A.L., Herald, T.J., Schmidt, K.A., Tilley, K.A., *J. Agric. Food Chem*. 2001, *49*, 4983-4989.

[13] Natarajan, S.S., Xu, C., Bae, H., Caperna, T.J., Garrett, W.M., J. Agric. *Food Chem*. 2006, *54*, 3114-3120.

[14] Sanchez, L., Perez, M.D., Puyol, P., Calvo, M., Brett. G., *J. AOAC Int*. 2002, *85*, 1390-1397.

[15] Fernandez Ocaña, M., Fraser, P., Patel, R.K.P., Halket, J.M., Bramley, P.M., *Rapid. Commun. Mass Spectrom*. 2007, *21*, 319-328.

[16] Fernandez Ocaña, M., Fraser, P., Patel, R.K.P., Halket, J.M., Bramley, P.M., *Anal. Chim. Acta* 2009, *634*, 75-82.

[17] Krishnan, H.B., Natarajan, S.S., Mahmoud, A.A., Nelson, R.L., *J. Agric. Food Chem*. 2007, *55*, 1839-1845.

[18] Brandon, D.L., Friedman, M., *J. Agric. Food Chem*. 2002, *50*, 6635-6642.

[19] Natarajan, S., Xu, C., Caperna, T.J., Garrett, W.M., *Anal. Biochem*. 2005, *342*, 214-220.

[20] Heras, J.M., Marina, M.L., Garcia, M.C., *J. Chromatogr. A* 2007, *1153*, 97-103.

[21] Saz, J.M., Marina, M.L., *J. Sep. Sci*. 2007, *30*, 431-451.

[22] Leitner, A., Castro-Rubio, F., Marina, M.L., Lindner, W., *J. Proteome Res*. 2006, *5*, 2424-2430.

[23] Garcia, M.C., Heras, J.M., Marina, M.L., *J. Sep. Sci*. 2007, *30*, 475-482.

[24] Horneffer, V., Foster, T.J., Velikov, K.P., *J. Agric. Food Chem*. 2007, *55*, 10505-10508.

[25] McDonald, W.H., Yates 3rd, J.R., *Dis. Markers*. 2002, *18*, 99-105.

[26] Storms, H.F., van der Heijden, R., Tjaden, U.R., Van der Greef, J., *Electrophoresis* 2004, *25*, 3461-3467.

[27] Kasicka, V., *Electrophoresis* 2008, *29*, 179-206.

[28] Stutz, H,. *Electrophoresis* 2005, *26*, 1254-1290.

[29] Herrero, M., Ibañez, E., Cifuentes, A., *Electrophoresis* 2008, *29*, 2148-2160.

[30] Klampfl, C.W., *Electrophoresis* 2009, *30*, S83-S91.

[31] Sandra K, Moshir M, D'hondt F, Tuytten R, Verleysen K, Kas K, François I, Sandra P., J. Chromatogr. B 2009, 877, 1019-1039.

[32] Simó, C., Barbas, C., Cifuentes, A., *Electrophoresis* 2005, *26*, 1306-1318.

[33] Gaspar, A., Englmann, M., Fekete, A., Harir, M., Schmitt-Kopplin, P., *Electrophoresis* 2008, *29*, 66-79.

[34] Garcia-Cañas, V., Gonzalez, R., Cifuentes, A., *J. Agric. Food Chem*. 2002, *50*, 4497-4502.

[35] Garcia, M.C., Garcia, B., Garcia-Ruiz, C., Gomez, A., Cifuentes, A., Marina, M.L., *Food Chem*. 2009, *4*, 1212-1217.

[36] Tsuruki, T., Kishi, K., Takahashi, M., Tanaka, M., Matsukawa, T., Yoshikawa, M., *FEBS Lett*. 2003, *540*, 206-210.

[37] Faris, R.J., Wang, H., Wang, T., *J. Agric. Food Chem*. 2008, *56*, 7146-7150.

[38] Endres, J.G., in: Endres, J.G. (Ed.), *Soy protein products. Characteristics, nutritional aspects, and utilization*, AOAC Press, Champaign, 2001, pp. 4-19.

[39] Garcia-Villalba, R., Leon, C., Dinelli, G., Segura-Carretero, A., Fernández-Gutierrez, A., Garcia-Cañas, V., Cifuentes, A., *J. Chromatogr. A* 2008, *1195*, 164-173.

Figure legends

Figure 1. SDS-PAGE profile of two protein extracts from soybean. Lane 1: Molecular mass marker. Lane 2: extraction with 7M urea, 2 M thiourea, 3% w/v CHAPS and 40 mM Tris base. Lane 3: extraction with ACN-water 80:20 (v/v).

Figure 2. CE-TOF MS BPE obtained for the different soybean protein hydrolysates (continuous line) and their blanks (dotted line) obtained using: A) heat treatment and B) reductive alkylation tratment, at different trypsin:protein ratios.

Figure 3. CE-TOF MS BPE of the digested SPI extract using different concentrations of formic acid: A) 1.5 M, B) 1.0 M, C) 0.5 M, and D) 0.25 M.

Figure 4. CE-TOF MS BPE of the digested SPI extract under optimum conditions. Bare silica capillary (50 μm id, 90 cm); BGE: 0.5 M formic acid; injection time: 20 s at 0.5 psi (34.5 mbar); separation voltage: 25 kV; sheath liquid: isopropanol-water (50:50, v/v) at a 3 μL/min flow rate; nebulizer gas: 0.4 bar; drying gas: 4 L/min N_2 at 200 ºC; MS is used in positive ion mode; scan: 50-3000 m/z.

Figure 5. Selected $[M+H]^+$ mass spectra and their corresponding deconvoluted mass spectra of A) peak A, B) peak C and C) peak G as marked in Figure 4. CE-TOF MS analysis conditions as in Figure 4.

Figure 6. CE-TOF MS base peak electropherogram (BPE) of the digested SPI extract and their corresponding extracted ion electropherogram (EIEs). CE-TOF MS analysis

conditions as in Figure 4.

Figure 7. CE-TOF MS base peak electropherogram (BPE) of the digested protein extract from conventional and transgenic soybean. CE-TOF MS analysis conditions as in Figure 4.