André M. Murad[1]
Gustavo H. M. F. Souza[2]
Jerusa S. Garcia[3]
Elíbio L. Rech[1]

[1]Embrapa Genetic Resources and Biotechnology, Laboratory of Gene Transfer, Parque Estação Biológica, PqEB, Brazil
[2]Waters Corporation, MS Applications Research and Development Laboratory, Alameda Tocantins, São Paulo, Brazil
[3]Alfenas Federal University, Institute of Exact Sciences, Alfenas, MG, Brazil

Received March 15, 2011
Revised July 11, 2011
Accepted July 12, 2011

Research Article

Detection and expression analysis of recombinant proteins in plant-derived complex mixtures using nanoUPLC-MS^E

The use of mass spectrometry to identify recombinant proteins that are expressed in total soluble proteins (TSPs) from plant extracts is necessary to accelerate further processing steps. For example, the method consists of TSP sample preparation and trypsin digestion prior to the preliminary characterization using nanoUPLC-MS^E analysis of the recombinant proteins that are expressed in TSP samples of transgenic soybean seeds. A TSP sample as small as 50 μg can be effectively analyzed. In this study, transgenic soybean seeds that expressed recombinant cancer testis antigen (CTAG) were used. The procedure covered 30% of the protein sequence and was quantified at 0.26 ng, which corresponded to 0.1% of the TSP sample. A comparative proteomic profile was generated by the comparison of a negative control and sample that showed a unique expression pattern of CTAG in a transgenic line. The experimental data from the TSP extraction, sample preparation and data analysis are discussed herein.

Keywords: Expression^E / Identity^E / NanoLC-MS^E / ProteinLynx global server / Soybean
DOI 10.1002/jssc.201100238

1 Introduction

The production of recombinant proteins is an important step in several academic, industrial and pharmaceutical processes. Several heterologous protein expression systems are available, including bacterial [1], mammalian cell culture [2] and plant [3, 4] systems. Although these systems constitute the main production systems, the search for novel methods to increase protein yield, facilitate manipulation and reduce cost is ongoing. Seeds are an advantageous production platform for recombinant proteins for several reasons: they can undergo long-term storage at ambient temperatures [5, 6], they can provide an appropriate biochemical environment for protein stability through the creation of specialized storage compartments [6, 7], they are not contaminated by human or animal pathogens [8], their desiccation characteristics prevent them from undergoing non-enzymatic hydrolysis or protease degradation [5, 8], and they do not carry the harmful substances that are present in some plant leaves, which is important for downstream processing [3, 8]. We recently produced several transgenic soybean plants that express important pharmaceutical molecules, such as proinsulin [6], human growth hormone (hGH) [9] and human coagulation factor IX (hFIX) [10], which demonstrate the versatility of this system. However, the production of these transgenic lines is extremely time consuming [11] and requires at least 150 days to obtain the first seeds and another three years to generate a homozygous line. During the early stages, there is little material available for recombinant protein purification. As a result, the detection, quantification and characterization of recombinant molecules rely mainly on the manipulation of the total soluble protein (TSP), which is a complex mixture that contains a low abundance of the protein of interest. Therefore, a method that detects, qualifies and quantifies recombinant proteins in TSP using $<1/4$ of a single seed mass (≈ 50 mg) is needed.

Typically, a recombinant protein is identified using Western blot analysis [12] and quantified using enzyme-linked immunosorbent assays (ELISAs) [13]. These methods are widely used because they are simple and relatively fast for identification and quantification, but they lack sensitive detection ability when small amounts of antigens are used, no antibody is available, or a false positive is found, and there is no way to verify the quality, amino acid sequence or post-translational modifications of the recombinant protein. Two dimensional electrophoresis (2-DE) was developed for proteomics [14, 15], and because of its association with MS, it has become the primary tool for the identification and characterization of complex plant mixtures [15, 16]. 2-DE

Correspondence: Elíbio L. Rech, Embrapa Genetic Resources and Biotechnology, Laboratory of Gene Transfer, Parque Estação Biológica, PqEB, Av. W5 Norte, Brasília, DF, 70770-917, Brazil
E-mail: rech@cenargen.embrapa.br
Fax: +55-61-3448-4694

Abbreviations: CTAG, cancer testis antigen 1; **PLGS**, ProteinLynx global server; **TSP**, total soluble protein

can also be used for the quantification and protein mapping of tissues [17], comparative proteomics [18, 19] and post-translational modification identification [20]. However, this technique requires a minimum sample amount, cannot detect low-abundance molecules, requires spot manipulations for successful identification of proteins [15], is performed mainly using peptide mass fingerprinting (PMF) [21, 22], and has difficulty analyzing proteins with similar mass and p*I* because they appear as a single spot. The combination of gel and liquid chromatography mass spectrometry (LC-MS) methods may result in the better identification of proteins in complex samples [23, 24] and may overcome the drawbacks of 2-DE. LC increases the low detection/resolution of complex mixtures on mass spectrometers (MS) [25]. Furthermore, the analysis of peptides or complex samples that have been digested by trypsin, commonly known as ''system samples'', is key for the detection of low-abundance proteins, but this technique has limitations in analyte dilution, the minimum amounts of complex protein mixtures that are needed to guarantee a satisfactory dynamic range and the detection of low-abundance proteins [14, 15, 25, 26].

Nano-scale liquid chromatography combined with mass spectrometry with data-independent acquisitions (nanoLC-MS^E) has several benefits for proteome analysis. Among these benefits are detection and linear sequence structural information at the femtomole level [12, 27]; small surface areas and minimal dead volumes, which minimize analyte losses due to surface adsorption; and low flow rates, which reduce analyte dilution. Low-abundance analytes can be separated with a high recovery rate when they are associated with a high dynamic range and a high-quality MS detection system [13]. Recently, the nanoLC-MS method has been used for the detection of differences in the expression of soybean plasma membrane proteins under osmotic stress [28], the regulation of stress identification genes induced by iron deficiency in tomatoes [29] and the detection of neuropeptides that are secreted by *Cancer borealis* [30], which demonstrates the capability and potential of this method.

Moreover, nanoLC-MS is an important tool for the characterization of post-translational modifications of proteins, such as the identification of N-terminal peptide modifications in the chloroplast proteome [31], the analysis of human protein oxidations that lead to functional reduction/annulation [32], and the characterization of phosphorylation patterns of several phosphatase splice variants that are expressed in a human cell line [33, 34]. Finally, quantification is also possible with the nanoLC-MS technique using labeling methods, such as the (^{18}O) labeling of peptides [35] and the iTRAQ™ method [36], which are based on relative quantification methods, such as the use of stochastic measurements of mass and intensity deviations for each ion that is detected [37] or the absolute quantification based on a constant ion current that is acquired with low and high collision energies into the mass spectrometer, called MS^E [38–41]. MS^E is a data-independent acquisition method that uses low and high collision energies with no precursor selection, which is different from other methods, such as data-dependent acquisition (DDA). The ion detection, clustering and normalization of data-independent, alternate scanning LC-MS^E data have been explained in detail elsewhere [39].

Here, we describe an easy-to-handle, label-free nanoUPLC-MS^E method that provides absolute quantification and allows the use of small samples for the detection, quantification and characterization of low-abundance recombinant proteins that are expressed in soybean seeds, specifically the immunogenic tumor NY-ESO-1 antigen (cancer testis antigen 1, CTAG) [42]. CTAG is a protein product of a gene located on the human X chromosome that comprises 180 amino acid residues (Fig. 2) with a mass of 18 kDa, a glycine-rich N-terminal region and an extremely hydrophobic C-terminal region that is so insoluble that it bears striking resemblance to a transmembrane domain [42, 43]. Because of the insolubility of the C-terminal region, the production and subsequent identification and characterization of this protein from TSP extracts present a challenge. Expression pattern analysis using RT-PCR has confirmed that CTAG expression is restricted to the testis. It is not present in other normal tissues, but it is found in several types of cancer, including bladder, breast and lung cancers [44]. The recombinant CTAG that is produced in *Escherichia coli* (*E. coli*) was the first to be evaluated in the clinical setting in a study that ranks among the most promising CTAG trials that have been published because of the favorable and broadly applicable immunological clinical results [42, 45]. However, the use of CTAG as a vaccine is viable only if it is coupled with a low-cost, scalable recombinant protein production system.

2 Materials and methods

2.1 Chemicals and reagents

Water was purified using a Milli-Q purification system (Millipore, France) to obtain a conductivity of <1.3 μS/cm, a

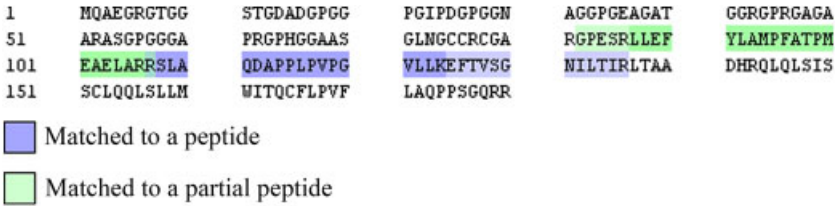

Figure 1. The CTAG amino acid sequence. The colored boxes indicate the peptides found using PLGS (Table 2), and the overlapping regions are indicated by changes in color.

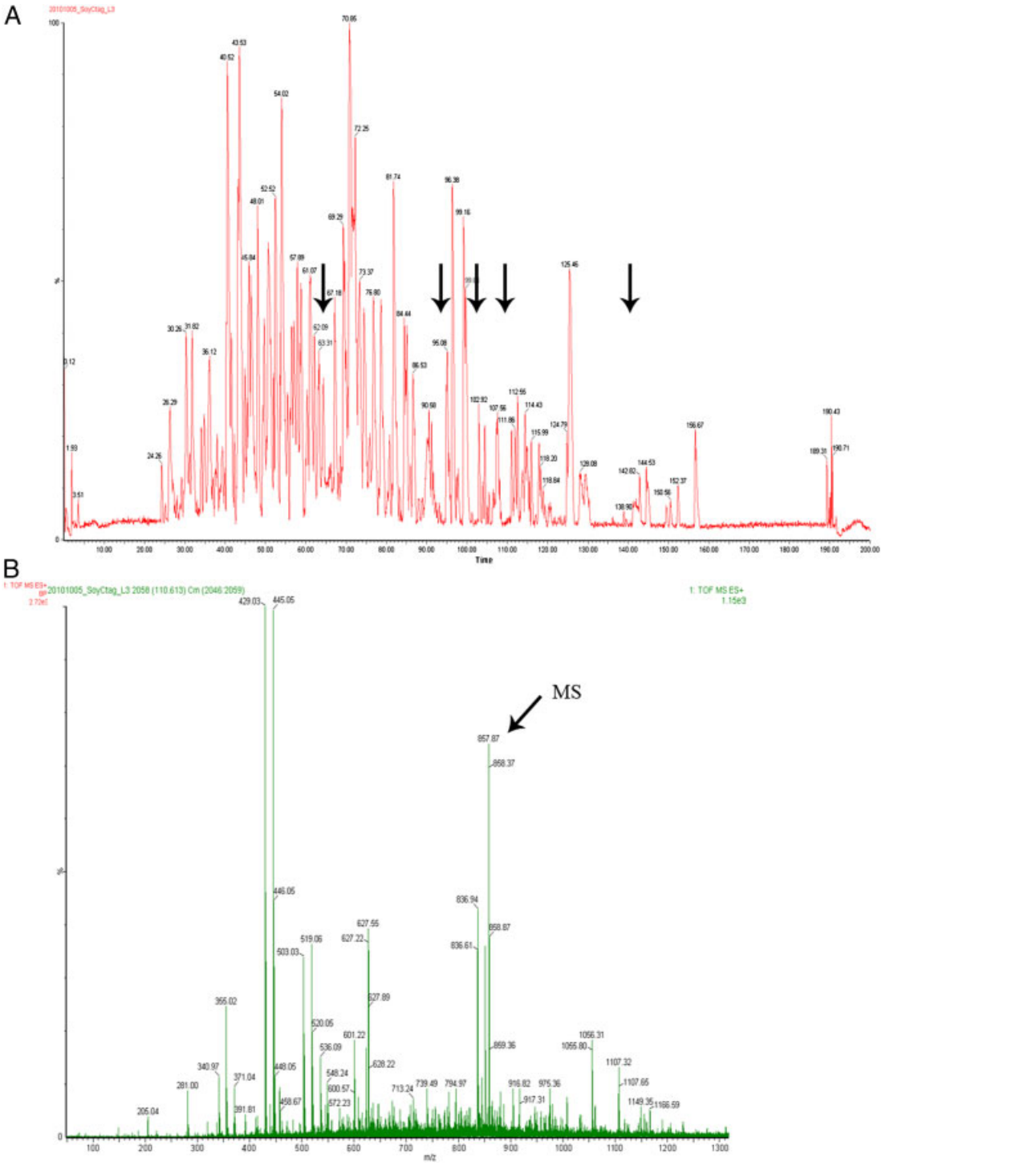

Figure 2. Experimental spectra results. (A) The nanoUPLC chromatogram of soybean CTAG lineage 3. The arrows indicate the eluted peptides corresponding to the CTAG-digested protein (Table 2). (B) MS spectra from the 110.31 min containing the $[M+2H]^{+2} = 857.87$ ion from the CTAG protein fragment. (C) Reconstructed product ion spectra from the $[M+2H]^{+2} = 857.87$ ion precursor. (D) Deconvoluted MS/MS spectra processed using PLGS and the peptide sequence corresponding to ion $[M+H]\ (1.1)^{+} = 1715.01$ from the trypsin digestion of CTAG.

total organic carbon (TOC) <2 ppb, and a semiconductor equivalent specification of 0.055 μS/cm (18.2 mΩ cm) at point-of-use at 25°C. Analytical-grade petroleum ether, 30–75°C, was purchased from J.T. Baker. Tris base (2-amino-2-[hydroxymethyl]-1,3-propanediol), KCl, DL-dithiothreitol (threo-1,4-dimercapto-2,3-butanediol; DTT), phenylmethanesulfonyl fluoride (PMSF) ≥98.5%, sodium dodecyl sulfate (SDS) for molecular biology ≥98.5%, acetone

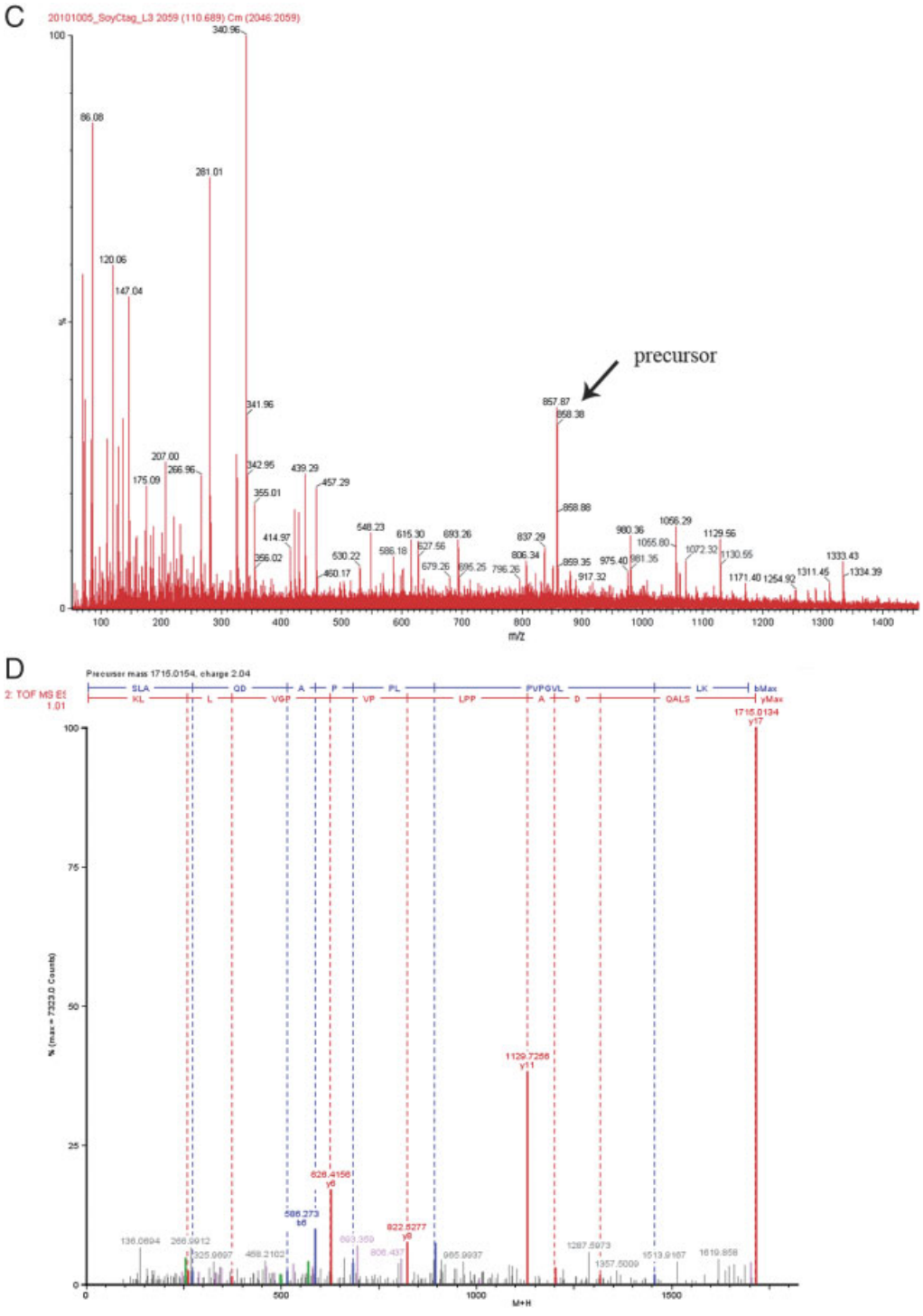

Figure 2. Continued.

CHROMASOLV® Plus ≥99.9%, ammonium bicarbonate ≥99.0%, iodoacetamide, trifluoroacetic acid, acetonitrile, formic acid and [Glu[1]]-fibrinopeptide B, human (GFP) were purchased from Sigma-Aldrich. MassPREP Protein Digestion Standard Alcohol Dehydrogenase (ADH) and *Rapi*-GEST™ SF were purchased from Waters (Waters, USA). Sequencing-grade trypsin was purchased from Promega.

2.2 TSP extraction from recombinant CTAG soybean seeds

Two samples, Soy CTAG L3 and Soy Cn negative seeds from the BR-16 cultivar, were used in this work. The soybean seeds were ground into a fine powder in a coffee grinder. A 100 mg sample of powder was weighed and placed into a

2 mL capped centrifuge tube. Petroleum ether (1 mL) was added, and the sample was slowly agitated for 15 min. The supernatant was discarded, and this step was repeated twice (2 ×). The petroleum ether was evaporated for 10 min, and 1 mL of 20 mM Tris-HCl, pH 8.3, 1.5 mM KCl, 10 mM DTT, 1 mM PMSF and 0.1% v/v SDS were added. The sample was slowly vortexed at room temperature for 10 min and centrifuged for 5 min at 5000 rpm at 4°C. The supernatant was then transferred to a new centrifuge tube. For each 200 μL of sample, 800 μL of cold acetone was added to the centrifuge tube. The sample was vortexed thoroughly and kept at −20°C for 1 h, with vortexing every 15 min. The sample was then centrifuged for 10 min at 13000 rpm. The supernatant was discarded, and the pellet was dried at room temperature for 30 min. The pellet was carefully dissolved in 500 μL of 50 mM NH_4HCO_3 and quantified using the Quant-iT™ Protein Assay Kit (Invitrogen, USA). The sample was finally diluted with 50 mM ammonium bicarbonate to a concentration of 1 μg/μL.

2.3 Sample preparation for nanoUPLC-MS^E acquisition

An aliquot of 50 μL of the 1 μg/μL sample was added to 10 μL of 50 mM ammonium bicarbonate in a microcentrifuge tube. Next, 25 μL of *Rapi*GEST™ (0.2% v/v) was added, and the sample was vortexed and incubated in a dry bath at 80°C for 15 min. The sample was briefly centrifuged, and 2.5 μL of 100 mM DTT was added. The sample was vortexed gently and incubated at 60°C for 30 min, followed by centrifugation. Iodoacetamide (2.5 μL of a 300 mM solution) was added, and the sample was vortexed slightly and incubated in the dark at room temperature for 30 min. Next, 10 μL of trypsin (with 400 μL of 50 mM ammonium bicarbonate added per 20 μg vial of trypsin) was added, and the sample was vortexed slightly. The sample was digested at 37°C in a dry bath overnight. To precipitate the *Rapi*GEST, 10 μL of a 5% TFA solution was added, and the sample was vortexed, incubated for 90 min at 37°C in a dry bath, and centrifuged at 14 000 rpm at 6°C for 30 min. The supernatant was transferred to a Waters Total Recovery vial (Waters), and 5 μL of ADH (with 1 mL of 3% acetonitrile and 0.1% formic acid) and 85 μL of a 3% acetonitrile, 0.1% formic acid solution was added. The final concentration of the protein was 250 ng/μL, and the final concentration of ADH was 25 fmol/μL. The final volume was 200 μL.

2.4 NanoLC-MS^E acquisition

Nanoscale LC separation of tryptic peptides from TSP was performed using a nanoACQUITY™ system (Waters) equipped with a Symmetry C18 5 μm, 5-mm × 300-μm precolumn and a nanoEase™ BEH130 C18 1.7 μm, 100 μm × 100 mm analytical reversed-phase column (Waters). The samples were initially transferred to the precolumn using an aqueous 0.1% formic acid solution with a flow rate of 15 μL/min for 1 min. Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile. The peptides were separated using a gradient of 3–40% mobile phase B for 200 min with a flow rate of 600 nL/min followed by a 10 min rinse with 90% of mobile phase B. The column was re-equilibrated to the initial conditions for 20 min. The column temperature was maintained at 35°C. The lock mass was delivered from the auxiliary pump of the nanoACQUITY pump using a constant flow rate of 150 nL/min at a concentration of 100 fmol of GFP to the reference sprayer of the NanoLock-Spray source of the mass spectrometer. All samples were analyzed in triplicate.

The tryptic peptides were analyzed using a Synapt HDMS™ mass spectrometer (Waters, Manchester, UK) with a hybrid quadrupole/ion mobility/orthogonal acceleration time-of-flight (*oa*-TOF) geometry. For all measurements, the mass spectrometer was operated in the ''V-mode'' of analysis with a typical resolving power of at least 10 000 full-width half-maximum (FWHM). All analyses were performed using a positive nanoelectrospray ion mode (nanoESI$^+$). The time-of-flight analyzer of the mass spectrometer was externally calibrated with GFP b+ and y+ ions from *m/z* 50 to 1990 with the data post-acquisition lock mass corrected using the GFP double charged precursor ion $[M+2H]^{2+} = 785.8426$. The reference sprayer was sampled at a frequency of 30 s. Exact mass retention time (EMRT) [46] nanoLC-MS^E data were collected in an alternating low-energy and elevated energy mode of acquisition. The continuum spectra acquisition time, in each mode, was 1.5 s with a 0.1 s interscan delay. In the low-energy MS mode, data were collected at constant collision energy of 3 eV. In

Table 1. Peptide sequences found using PLGS for CTAG protein

Precursor MH+(Da)	Charge state	MH+Error (Da)	Score	Start	End	Sequence	Modifications	Retention time (min)	Intensity
1715.0154	2	0.0032	7.8859	108	124	(R)SLAQDAPPLPVPGVLLK(E)		110.31	19 656
1349.7391	2	0.0019	7.2048	125	136	(K)EFTVSGNILTIR(L)		94.51	7191
1871.1141	3	−0.0038	7.0925	107	124	(R)RSLAQDAPPLPVPGVLLK(E)		102.18	2493
2485.308	3	0.0177	6.1839	87	107	(R)LLEFYLAMPFATPMEAELARR(S)	Oxidation M (8)	141.81	2994
2855.4158	4	0.0052	6.0799	82	106	(R)GPESRLLEFYLAMPFATPMEAELAR(R)	Oxidation M (13)	63.48	2744

Table 2. Proteins identified using PLGS in the Soy CTAG L3 transgenic soybean line

Entry	Description	MW (Da)	p*I* (pH)	PLGS score	Amount (ng)	% of TSP
P78358	Cancer testis antigen 1	17 981	8.47	2886.38	0.26	0.11
O22120	α-Subunit of β-conglycinin Fragment OS Glycine max PE 2 SV 2	63 126	4.72	51 090.21	65.47	28.43
C6T488	Putative uncharacterized protein OS Glycine max PE 2 SV 1	24 103	5.13	49 332.05	0	0
P04776	Glycinin G1 OS Glycine max GN GY1 PE 1 SV 2	55 671	5.82	34 638.61	28.68	12.45
P19594	2S albumin OS Glycine max PE 1 SV 2	18 447	5.01	26 866.75	6.98	3.03
Q549Z4	Proglycinin A2B1 OS Glycine max PE 2 SV 1	54 356	5.29	26 163.09	9.12	3.96
P04405	Glycinin G2 OS Glycine max GN Gy2 PE 1 SV 2	54 356	5.29	26 155.1	16.83	7.31
C6TKH0	Putative uncharacterized protein OS Glycine max PE 2 SV 1	31 640	6.41	25 943.07	4.40	1.91
B3TDK4	Lipoxygenase OS Glycine max PE 3 SV 1	94 352	5.87	23 234.38	11.22	4.87
P08170	Seed lipoxygenase 1 OS Glycine max GN LOX1 1 PE 1 SV 2	94 310	5.93	22 866.77	0	0
P01063	Bowman Birk-type proteinase inhibitor C II OS Glycine max PE 1 SV 2	9194	4.37	19 673.12	2.77	1.20
P01064	Bowman Birk-type proteinase inhibitor D II OS Glycine max PE 1 SV 2	9460	4.66	18 789.71	1.43	0.62
P24337	Hydrophobic seed protein OS Glycine max PE 1 SV 1	8353	6.04	17 254.79	0.54	0.23
Q39805	Dehydrin-like protein OS Glycine max PE 2 SV 1	23 703	6.084	16 428.98	5.44	2.36
Q7GC77	Glycinin A3B4 subunit OS Glycine max PE 1 SV 1	58 151	5.41	14 016.68	0.31	0.13
Q852U4	Glycinin A1bB2 784 OS Glycine max PE 2 SV 1	54 264	5.94	12 395.54	0.65	0.28
Q852U5	Glycinin A1bB2 445 OS Glycine max PE 2 SV 1	54 183	5.77	12 393.59	0.49	0.21
P05046	Lectin OS Glycine max GN LE1 PE 1 SV 1	30 908	5.59	11 981.52	8.64	3.75
C6T9Z5	Putative uncharacterized protein OS Glycine max PE 2 SV 1	42 796	6.29	11 699.31	1.09	0.47
C6TDF5	Putative uncharacterized protein OS Glycine max PE 2 SV 1	41 854	6.99	11 311.33	0.50	0.22
Q9SEK9	Seed maturation protein PM25 OS Glycine max GN PM25 PE 2 SV 1	25 713	4.78	9964.62	1.03	0.45
Q9SEK8	Seed maturation protein PM26 OS Glycine max GN PM26 PE 2 SV 1	26 087	4.63	9770.58	0.97	0.42
Q9XET1	Seed maturation protein PM31 OS Glycine max GN PM31 PE 2 SV 1	17 735	6.10	9168.43	1.50	0.65
Q9SEL0	Seed maturation protein PM24 OS Glycine max GN PM24 PE 2 SV 1	26 824	4.97	8024.35	0.87	0.38
Q9XER5	Seed maturation protein PM22 OS Glycine max GN PM22 PE 2 SV 1	16 677	4.96	7963.37	0.61	0.26
Q9LLQ6	Seed maturation protein PM34 OS Glycine max GN PM34 PE 2 SV 1	31 746	6.68	7863.56	0.43	0.18
C6T1Q7	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 812	5.95	7729.99	1.31	0.57
C6T588	Putative uncharacterized protein OS Glycine max PE 2 SV 1	16 750	4.53	7076.10	0.60	0.26
Q9AVK8	Allergen Gly m Bd 28K Fragment OS Glycine max PE 2 SV 1	52 608	5.65	5503.08	2.86	1.24
Q2I0H4	Glyceraldehyde 3 phosphate dehydrogenase OS Glycine max GN GAPC1 PE 2 SV 1	36 741	6.84	5307.64	1.46	0.63
Q9XET0	Putative uncharacterized protein OS Glycine max GN PM30 PE 2 SV 1	15 088	9.42	5145.05	0.76	0.33
C6TBB3	Putative uncharacterized protein OS Glycine max PE 4 SV 1	12 337	5.38	4899.04	0.12	0.05
P93165	Em protein OS Glycine max PE 4 SV 1	11 484	5.35	4895.39	0.08	0.03
Q04672	Sucrose-binding protein OS Glycine max GN SBP PE 1 SV 1	60 484	6.42	4608.47	3.36	1.46
C6SVM2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 367	9.48	4510.81	0.52	0.22
Q07CZ3	Glyceraldehyde 3 dehydrogenase C subunit OS Glycine max PE 2 SV 1	36 701	6.84	4077.10	0	0
C6SWV3	Putative uncharacterized protein OS Glycine max PE 2 SV 1	27 618	5.66	3894.10	0.71	0.31
C6TB70	Putative uncharacterized protein OS Glycine max PE 2 SV 1	24 404	6.50	3485.42	0.73	0.31
Q9XES8	Seed maturation protein PM28 OS Glycine max GN PM28 PE 4 SV 1	9506	4.46	3193.13	0.18	0.08
C6T0L2	Putative uncharacterized protein OS Glycine max PE 4 SV 1	11 134	6.37	3078.30	1.15	0.49
Q38IW8	Triosephosphate isomerase OS Glycine max PE 2 SV 1	27 187	5.81	2909.14	0.20	0.08
Q9SWB2	Seed maturation protein PM41 OS Glycine max GN PM41 PE 4 SV 1	8172	4.66	2896.55	0.21	0.091
Q42795	β-Amylase OS Glycine max PE 1 SV 1	56 036	5.18	2892.25	2.01	0.87
Q39871	Late embryogenesis abundant protein OS Glycine max GN MP2 PE 2 SV 1	50 613	6.29	2760.77	4.41	1.91
C6T0B5	Putative uncharacterized protein OS Glycine max PE 2 SV 1	13 998	5.69	2535.70	0.43	0.19
C6SVR5	Putative uncharacterized protein OS Glycine max PE 2 SV 1	23 888	5.63	2489.41	0.42	0.18
P00330	ALCOHOL DEHYDROGENASE I EC 1 1 1 1	36 668	6.27	2351.86	0.91	0.39
C6SZ11	Putative uncharacterized protein OS Glycine max PE 2 SV 1	27 031	6.45	2284.27	0.54	0.23
O64458	Gly m Bd 30K allergen OS Glycine max GN P34 PE 2 SV 1	42 730	5.56	2154.71	3.44	1.49
C6TD82	Putative uncharacterized protein OS Glycine max PE 2 SV 1	31 058	7.50	2149.04	0.04	0.02
C6TCF1	Putative uncharacterized protein OS Glycine max PE 2 SV 1	27 781	5.09	2055.68	0.33	0.14
C6TB67	Putative uncharacterized protein OS Glycine max PE 2 SV 1	22 971	7.71	1790.65	0.33	0.14
C6EVF9	Elongation factor 1 α-OS Glycine max GN EF 1A PE 2 SV 1	49 365	9.23	1653.60	1.33	0.58
C6T072	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 442	5.26	1651.26	0.31	0.13
C6SWE0	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 355	5.26	1615.87	0.13	0.06
P26413	Heat shock 70 kDa protein OS Glycine max GN HSP70 PE 3 SV 1	70 835	5.18	1481.25	1.77	0.76
Q6RIB6	Malate dehydrogenase OS Glycine max PE 2 SV 1	35 504	6.34	1466.79	0.45	0.19

Table 2. Continued

Entry	Description	MW (Da)	p*I* (pH)	PLGS score	Amount (ng)	% of TSP
C6TK76	Putative uncharacterized protein OS Glycine max PE 2 SV 1	41 507	5.15	1249.91	0.13	0.06
C6TGM9	Putative uncharacterized protein OS Glycine max PE 2 SV 1	22 317	5.92	1194.81	0.11	0.05
A1KR24	Dehydrin OS Glycine max GN LEA 2 D11 PE 3 SV 1	25 369	6.11	1148.55	0.80	0.35
C6T920	Phosphoglycerate kinase Fragment OS Glycine max PE 2 SV 1	25 296	9.71	1114.19	0.24	0.10
Q84V19	Sucrose-binding protein 2 OS Glycine max GN SBP2 PE 2 SV 1	55 740	6.10	1069.19	0.30	0.13
Q9SP11	Sucrose-binding protein homolog S 64 OS Glycine max GN SBP PE 2 SV 1	55 799	6.31	1005.86	0.23	0.10
Q96450	14-3-3-like protein A OS Glycine max GN GF14A PE 2 SV 1	29 030	4.49	910.00	0.40	0.17
C6T9C2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	34 556	5.80	842.04	0.85	0.36
Q71EW8	Methionine synthase OS Glycine max PE 2 SV 1	84 229	5.88	833.61	1.34	0.58
C6K8D1	Seed biotinylated protein 68 kDa isoform OS Glycine max PE 2 SV 1	67 906	6.14	781.99	6.58	2.86
C6SZX7	Glutathione peroxidase OS Glycine max PE 2 SV 1	18 491	6.94	737.67	0.16	0.07
C6T1V2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 729	6.39	727.80	0.18	0.08
C6TNU2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	47 497	5.66	684.35	0.05	0.02
P27066	Ribulose bisphosphate carboxylase large chain OS Glycine max GN rbcL PE 1 SV 3	52 576	5.97	675.63	0.54	0.23
C6TB98	Putative uncharacterized protein OS Glycine max PE 2 SV 1	33 906	5.50	664.43	0.38	0.16
C6T8D8	Fructose bisphosphate aldolase Fragment OS Glycine max PE 2 SV 1	28 937	7.11	633.63	0.31	0.13
C6T4R9	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 656	10.15	584.85	0.32	0.14
C6SZN7	Putative uncharacterized protein OS Glycine max PE 4 SV 1	12 980	5.14	576.56	0.16	0.06
C6TLT3	Putative uncharacterized protein OS Glycine max PE 2 SV 1	29 708	10.23	557.49	0.09	0.04
C6TMG1	Fructose bisphosphate aldolase OS Glycine max PE 2 SV 1	38 315	7.34	534.53	0.30	0.13
C6T049	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 988	5.38	529.73	0.11	0.05
Q6RIB7	Enolase OS Glycine max PE 2 SV 1	47 689	5.14	518.47	0.53	0.23
C6T4Z6	Putative uncharacterized protein OS Glycine max PE 2 SV 1	15 883	10.54	502.43	0.04	0.01
C6SVT0	Putative uncharacterized protein OS Glycine max PE 2 SV 1	18 011	6.93	460.71	0.09	0.03
Q39839	Nucleoside diphosphate kinase 1 OS Glycine max PE 2 SV 1	16 432	5.88	458.52	0	0
C6SYU0	Putative uncharacterized protein OS Glycine max PE 2 SV 1	18 273	8.62	422.07	0.15	0.06
C6SZN6	Putative uncharacterized protein OS Glycine max PE 2 SV 1	17 935	5.50	386.40	0.04	0.02
C6TG05	Putative uncharacterized protein OS Glycine max PE 2 SV 1	40 355	6.52	381.64	0.32	0.13
C6K8D0	Trypsin inhibitor 26 kDa isoform OS Glycine max PE 2 SV 1	25 930	8.40	377.51	0.27	0.11
C6T1R3	Putative uncharacterized protein OS Glycine max PE 2 SV 1	25 133	5.96	362.56	0.18	0.08
Q9SPB8	Malate dehydrogenase OS Glycine max GN Mdh1 PE 3 SV 1	36 119	8.22	330.78	0.20	0.08
Q0GJJ9	ACP thioesterase protein Fragment OS Glycine max GN FATB1b PE 4 SV 1	22 492	5.48	318.69	0.81	0.35
Q9SWB4	Poly ADP ribose polymerase 3 OS Glycine max GN PARP3 PE 2 SV 1	91 630	5.26	316.88	0.83	0.36
C6TNI8	Putative uncharacterized protein Fragment OS Glycine max PE 2 SV 1	20 580	5.56	313.77	0.32	0.14
P29530	P24 oleosin isoform A OS Glycine max PE 2 SV 2	23 487	9.05	298.62	0.30	0.13
C6SXU0	Putative uncharacterized protein OS Glycine max PE 2 SV 1	27 721	6.81	295.30	0.23	0.10
C6SW79	40S ribosomal protein S12 OS Glycine max PE 2 SV 1	14 788	5.17	249.56	0.09	0.04
P28551	Tubulin β-chain Fragment OS Glycine max GN TUBB PE 2 SV 2	45 721	5.55	246.58	0.09	0.04
C6T7U2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	51 462	5.25	242.40	0.57	0.25
Q39801	51 kDa seed maturation protein OS Glycine max PE 2 SV 1	50 951	6.74	239.27	0.24	0.10
C6SVF1	Putative uncharacterized protein OS Glycine max PE 2 SV 1	15 968	6.15	237.95	0.04	0.01
C6TGA6	Putative uncharacterized protein OS Glycine max PE 2 SV 1	34 167	4.80	227.22	0.09	0.03
C6SXS9	Putative uncharacterized protein OS Glycine max PE 2 SV 1	23 473	5.77	219.11	0.14	0.06
Q8RVH5	Basic 7S globulin 2 OS Glycine max PE 1 SV 1	47 174	8.17	205.53	0	0
C6TBB8	Putative uncharacterized protein OS Glycine max PE 2 SV 1	13 300	8.96	190.49	0.19	0.08
B1Q2X4	Protein disulfide isomerase OS Glycine max GN PDIL 1 PE 3 SV 1	58 554	4.95	190.13	0.52	0.22
B0M1A9	Peroxisomal 3 ketoacyl CoA thiolase OS Glycine max PE 2 SV 1	48 585	7.70	189.25	0.26	0.11
B1ACD5	Kunitz trypsin protease inhibitor OS Glycine max PE 2 SV 1	22 661	5.06	177.80	0	0
C6TNU3	Putative uncharacterized protein OS Glycine max PE 2 SV 1	37 948	5.86	174.80	0.15	0.06
P54774	Cell division cycle protein 48 homolog OS Glycine max GN CDC48 PE 2 SV 1	89 713	5.00	169.92	0.82	0.35
C6T9X5	Putative uncharacterized protein OS Glycine max PE 2 SV 1	43 553	5.77	169.90	0.22	0.09
C6SXR4	Putative uncharacterized protein OS Glycine max PE 2 SV 1	14 832	11.30	166.68	0.05	0.02
C6TCR6	Putative uncharacterized protein OS Glycine max PE 2 SV 1	36 175	4.66	163.95	0.18	0.07
C6T262	Putative uncharacterized protein OS Glycine max PE 2 SV 1	22 373	7.29	163.30	0.31	0.13
C6TN03	Putative uncharacterized protein OS Glycine max PE 2 SV 1	29 982	10.63	144.91	0.66	0.29
C6TL46	Putative uncharacterized protein OS Glycine max PE 2 SV 1	29 582	5.75	142.51	0.77	0.33
A4ZGT5	Transcription factor bZIP129 Fragment OS Glycine max GN bZIP129 PE 2 SV 1	20 603	9.60	142.37	0.78	0.34
Q9SPJ6	Maturation protein pPM32 OS Glycine max GN PM32 PE 2 SV 1	18 871	5.31	135.95	0.24	0.10

Table 2. Continued

Entry	Description	MW (Da)	p*I* (pH)	PLGS score	Amount (ng)	% of TSP
C6TG88	Putative uncharacterized protein Fragment OS Glycine max PE 2 SV 1	17 907	4.58	131.27	1.35	0.58
C6ZRP9	Pti1 kinase-like protein OS Glycine max PE 2 SV 1	34 932	8.93	130.00	0.10	0.04
Q7XAC5	Embryo-specific urease OS Glycine max PE 2 SV 1	90 099	5.61	128.19	0.45	0.19
C6TJD3	Putative uncharacterized protein OS Glycine max PE 2 SV 1	35 726	7.61	122.80	0.15	0.06
O22518	40S ribosomal protein SA OS Glycine max PE 2 SV 1	33 885	4.90	122.27	0.23	0.10
C6TGJ9	Putative uncharacterized protein OS Glycine max PE 2 SV 1	28 237	11.13	117.53	0.04	0.01
Q0PJB9	MYB transcription factor MYB131 Fragment OS Glycine max GN MYB131 PE 2 SV 1	36 138	9.08	111.34	0.70	0.30
Q8L7J4	Pyruvate kinase OS Glycine max PE 2 SV 1	55 280	7.08	109.98	0.22	0.09
C6T520	Putative uncharacterized protein OS Glycine max PE 2 SV 1	18 088	5.80	109.35	0.20	0.08
C6T6B2	Putative uncharacterized protein OS Glycine max PE 2 SV 1	19 918	5.16	104.24	0.15	0.06
C6SY64	Proteasome subunit β-type OS Glycine max PE 2 SV 1	24 533	7.04	103.56	0	0
C6T470	Putative uncharacterized protein OS Glycine max PE 2 SV 1	27 220	4.52	101.06	0.07	0.03
B0M1A8	Peroxisomal aminotransferase Fragment OS Glycine max PE 2 SV 1	31 458	6.0	100.99	0.37	0.16

the elevated energy MS mode, the collision energy was ramped from 12 to 45 eV during each 1.5 s spectrum. The radiofrequency that was applied to the quadrupole mass analyzer was adjusted so that the ions from *m/z* 50 to 2000 were efficiently transmitted, which ensured that any ions less than *m/z* 50 that were observed in the LC-MS data only arose from dissociations in the TRAP T-wave collision cell.

2.5 Data processing and protein identification

The MS data that were obtained from the LC-MS^E were processed and searched using the ProteinLynx Global Server (PLGS) version 2.4 (Waters). Proteins were identified using the software's embedded ion accounting algorithm and a search of the Glycine max database with MassPREP digestion standards (MPDS) UniProtKB/Swiss-Prot sequences (Phosphorylase – P00489 – PHS2_RABIT, Bovine Hemoglobin – P02070 – HBB_BOVIN, ADH – P00330 – ADH1_YEAST, BSA – P02769 – ALBU_BOVIN) and CTAG (UniProtKB/Swiss-Prot–P78358–Cancer_testis_1) that were appended to the database. Identifications and quantitative data packaging were generated using dedicated algorithms [38, 47] and a search against a species-specific Uniprot database. The ion detection, clustering, and log-scale parametric normalizations were performed in PLGS with an Expression^E license installed. The intensity measurements were typically adjusted on these components, i.e. deisotoped and charge state-reduced EMRTs that were replicated throughout the complete experiment for the analysis at the EMRT cluster level. The fixed modification of carbamidomethyl-C was specified, and the variable modifications included were acetyl N-terminus, deamidation N, deamidation Q and oxidation M. Components were typically clustered with a 10 ppm mass precision and a 0.25-min time tolerance against the database-generated theoretical peptide ion masses with a minimum of one matched peptide. The alignment of elevated energy ions with low-energy precursor peptide ions was conducted with an approximate precision of 0.05 min. One missed cleavage site was allowed. The precursor and fragment ion tolerances were determined automatically. The protein identification criteria also included the detection of at least three fragment ions per peptide and the determination of at least one peptide per protein, and the identification of the protein was allowed with a maximum 4% false-positive discovery rate in at least three technical repeatability injections. Using protein identification replication as a filter, the false-positive rate was minimized because false-positive protein identifications, i.e. chemical noise, have a random nature and do not tend to replicate across injections. For the analysis of protein identification and quantification level, the observed intensity measurements were normalized to the intensity measurement of the identified peptides of the digested internal standard.

3 Results and discussion

3.1 Detection of recombinant CTAG protein

The CTAG amino acid sequence is shown in Fig. 1. The sample preparation from the TSP step to the nanoUPLC procedure is critical for a successful identification. The use of high-purity water and reagents is recommended because of the sensitivity of the technique. With the low peptide dilution that is provided by nanoUPLC, each peptide is carried into the mass spectrometer almost individually, which allows for the production of precursor and fragment spectra from almost every peptide in the sample. At this separation power, the number of spread chromatography peaks is reduced to a minimum. When nanoACQUITY is associated with MS^E acquisitions [39], the ion current is continuous, both precursors and product ions are acquired in parallel, the chromatographic peaks are sharpened as more points per peak are obtained, and there is high

Table 3. Expression analysis comparison of control (Soy Cn) and expressed (Soy CTAG L3) samples. The first values correspond to the ratio values from the samples. The log(*e*) of the ratio and its standard deviation are shown in parentheses. The *p*-values corresponding to values from 0 to 1, in which 0–0.05 was considered down-regulated and 0.95–1.00 was considered up-regulated, are shown in brackets

Accession	Description	Score	Soy Cn: Soy CTAG L3
CTG1B_HUMAN	Cancer testis antigen 1	2886.4	Soy CTAG L3 Unique
Q42447_SOYBN	Maturation protein OS Glycine max GN MAT1 PE 2 SV 1	666.94	0.08 (−2.51 ± 0.73) [0]
OLEO1_SOYBN	P24 oleosin isoform A OS Glycine max PE 2 SV 2	298.62	0.09 (−2.46 ± 1.25) [0]
C6T8B0_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	372.96	0.23 (−1.45 ± 0.69) [0]
O22120_SOYBN	α-Subunit of β conglycinin Fragment OS Glycine max PE 2 SV 2	51 090.21	0.24 (−1.41 ± 0.16) [0]
Q42780_SOYBN	Lipoxygenase OS Glycine max GN lox7 PE 2 SV 1	2381.86	0.25 (−1.4 ± 0.67) [0]
Q9AVK8_SOYBN	Allergen Gly m Bd 28K Fragment OS Glycine max PE 2 SV 1	5503.09	0.25 (−1.38 ± 0.36) [0]
Q9LLQ6_SOYBN	Seed maturation protein PM34 OS Glycine max GN PM34 PE 2 SV 1	7863.56	0.38 (−0.97 ± 0.48) [0]
C6SXV3_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	269.85	0.39 (−0.95 ± 0.68) [0]
C6TN36_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	2764.85	0.41 (−0.88 ± 0.24) [0]
C6SZX7_SOYBN	Glutathione peroxidase OS Glycine max PE 2 SV 1	737.68	0.42 (−0.86 ± 0.61) [0.02]
Q4U3W3_SOYBN	β-Amylase OS Glycine max PE 3 SV 1	1151.52	0.44 (−0.81 ± 0.21) [0]
C6TDF5_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	11 311.33	0.45 (−0.8 ± 0.27) [0]
C6K8D0_SOYBN	Trypsin inhibitor 26 kDa isoform OS Glycine max PE 2 SV 1	377.51	0.54 (−0.62 ± 0.4) [0]
C6T0L2_SOYBN	Putative uncharacterized protein OS Glycine max PE 4 SV 1	3078.31	0.55 (−0.59 ± 0.36) [0.01]
IBBC2_SOYBN	Bowman Birk-type proteinase inhibitor C II OS Glycine max PE 1 SV 2	19 673.12	0.57 (−0.57 ± 0.09) [0]
C6TA60_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	258.09	0.64 (−0.44 ± 0.54) [0.1]
C6T920_SOYBN	Phosphoglycerate kinase Fragment OS Glycine max PE 2 SV 1	1114.2	0.64 (−0.44 ± 0.37) [0.02]
C6TNA8_SOYBN	Malate dehydrogenase OS Glycine max PE 2 SV 1	722.11	0.7 (−0.36 ± 1.38) [0.19]
ADH1_YEAST	ALCOHOL DEHYDROGENASE I EC 1 1 1 1	2351.86	0.7 (−0.36 ± 0.19) [0]
2SS_SOYBN	2S albumin OS Glycine max PE 1 SV 2	26 866.75	0.73 (−0.31 ± 0.1) [0]
Q39871_SOYBN	Late embryogenesis abundant protein OS Glycine max GN MP2 PE 2 SV 1	2760.78	0.74 (−0.3 ± 0.13) [0]
OLEO2_SOYBN	P24 oleosin isoform B OS Glycine max PE 2 SV 1	150.05	0.75 (−0.29 ± 0.46) [0.11]
C6TMG1_SOYBN	Fructose bisphosphate aldolase OS Glycine max PE 2 SV 1	534.53	0.75 (−0.29 ± 0.5) [0.16]
Q71EW8_SOYBN	Methionine synthase OS Glycine max PE 2 SV 1	833.61	0.76 (−0.27 ± 0.26) [0.03]
Q6RIB6_SOYBN	Malate dehydrogenase OS Glycine max PE 2 SV 1	1466.79	0.77 (−0.26 ± 0.27) [0.04]
PARP3_SOYBN	Poly ADP ribose polymerase 3 OS Glycine max GN PARP3 PE 2 SV 1	316.88	0.79 (−0.24 ± 0.36) [0.1]
Q39801_SOYBN	51 kDa seed maturation protein OS Glycine max PE 2 SV 1	239.28	0.79 (−0.24 ± 0.51) [0.19]
C6TG05_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	381.64	0.8 (−0.22 ± 0.38) [0.13]
RBL_SOYBN	Ribulose bisphosphate carboxylase large chain OS Glycine max GN rbcL PE 1 SV 3	675.63	0.8 (−0.22 ± 0.33) [0.09]
Q9SPJ6_SOYBN	Maturation protein pPM32 OS Glycine max GN PM32 PE 2 SV 1	135.96	0.8 (−0.22 ± 0.59) [0.29]
Q9ZT38_SOYBN	Alcohol dehydrogenase Fragment OS Glycine max GN Adh 2 PE 3 SV 1	1260.14	0.8 (−0.22 ± 0.46) [0.19]
IBBD2_SOYBN	Bowman Birk-type proteinase inhibitor D II OS Glycine max PE 1 SV 2	18 789.71	0.81 (−0.21 ± 0.42) [0.16]
Q9SEK8_SOYBN	Seed maturation protein PM26 OS Glycine max GN PM26 PE 2 SV 1	9770.59	0.83 (−0.19 ± 0.18) [0.04]
Q38IW8_SOYBN	Triosephosphate isomerase OS Glycine max PE 2 SV 1	2909.14	0.83 (−0.19 ± 0.45) [0.24]
C6T8W9_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	7690.69	0.83 (−0.19 ± 0.14) [0.02]
C6SZN7_SOYBN	Putative uncharacterized protein OS Glycine max PE 4 SV 1	576.57	0.84 (−0.18 ± 0.74) [0.28]
Q9SEL0_SOYBN	Seed maturation protein PM24 OS Glycine max GN PM24 PE 2 SV 1	8024.35	0.84 (−0.18 ± 0.18) [0.02]
Q70EM0_SOYBN	Dehydrin OS Glycine max GN lea D 11 PE 3 SV 1	16 419.81	0.84 (−0.18 ± 0.08) [0]
C6SWW4_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	28 100.36	0.84 (−0.17 ± 0.06) [0]
B3TDK6_SOYBN	Lipoxygenase OS Glycine max PE 3 SV 1	22 782.34	0.84 (−0.17 ± 0.05) [0]
C6TLT3_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	557.49	0.85 (−0.16 ± 0.81) [0.34]
C6SWV3_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	3894.11	0.86 (−0.15 ± 0.21) [0.13]
Q9SWB2_SOYBN	Seed maturation protein PM41 OS Glycine max GN PM41 PE 4 SV 1	2896.55	0.87 (−0.14 ± 0.29) [0.19]
Q9SEK9_SOYBN	Seed maturation protein PM25 OS Glycine max GN PM25 PE 2 SV 1	9964.62	0.87 (−0.14 ± 0.13) [0.03]
C6T7U2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	242.41	0.88 (−0.13 ± 0.31) [0.21]
Q39873_SOYBN	Lea protein OS Glycine max PE 2 SV 1	198.82	0.88 (−0.13 ± 0.52) [0.31]
KTI2_SOYBN	Kunitz-type trypsin inhibitor KTI2 OS Glycine max GN KTI2 PE 2 SV 1	3415.15	0.89 (−0.12 ± 0.25) [0.24]
B3TDK5_SOYBN	Lipoxygenase OS Glycine max PE 3 SV 1	21 743.68	0.9 (−0.11 ± 0.07) [0]
Q9XET1_SOYBN	Seed maturation protein PM31 OS Glycine max GN PM31 PE 2 SV 1	9168.43	0.9 (−0.11 ± 0.13) [0.07]
C6SZ11_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	2284.28	0.9 (−0.11 ± 0.45) [0.37]
B3TDK4_SOYBN	Lipoxygenase OS Glycine max PE 3 SV 1	23 234.38	0.9 (−0.1 ± 0.08) [0.01]
Q43440_SOYBN	Lipoxygenase OS Glycine max GN vlxC PE 2 SV 1	7075.76	0.91 (−0.09 ± 0.5) [0.35]
C6TMK3_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	202.81	0.92 (−0.08 ± 1.16) [0.45]
LOXX_SOYBN	Seed lipoxygenase OS Glycine max GN LOX1 4 PE 1 SV 1	6111.07	0.93 (−0.07 ± 0.34) [0.41]

Table 3. Continued

Accession	Description	Score	Soy Cn: Soy CTAG L3
C6TKH0_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	25 943.07	0.93 (−0.07 ± 0.08) [0.04]
C6SZK3_SOYBN	Glutathione peroxidase OS Glycine max PE 2 SV 1	737.65	0.93 (−0.07 ± 0.72) [0.46]
Q587K1_SOYBN	BiP OS Glycine max GN Gm bip PE 2 SV 1	744.24	0.94 (−0.06 ± 0.43) [0.43]
C6T0B5_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	2535.7	0.94 (−0.06 ± 0.22) [0.25]
C6TB98_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	664.43	0.95 (−0.05 ± 0.49) [0.41]
A7LCD5_SOYBN	Lipoxygenase OS Glycine max GN LOX9 PE 2 SV 1	791.73	0.95 (−0.05 ± 0.34) [0.45]
C6TBB3_SOYBN	Putative uncharacterized protein OS Glycine max PE 4 SV 1	4899.04	0.95 (−0.05 ± 0.69) [0.44]
Q9XES8_SOYBN	Seed maturation protein PM28 OS Glycine max GN PM28 PE 4 SV 1	3193.13	0.96 (−0.04 ± 0.25) [0.35]
C6T1Q7_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	7730	0.96 (−0.04 ± 0.13) [0.26]
C6SVM2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	4510.81	0.96 (−0.04 ± 0.24) [0.39]
Q948Y0_SOYBN	β-Conglycinin α prime subunit OS Glycine max PE 2 SV 1	23 517.56	0.96 (−0.04 ± 0.06) [0.13]
C6T9C2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	842.05	0.98 (−0.02 ± 0.2) [0.44]
HSP70_SOYBN	Heat shock 70 kDa protein OS Glycine max GN HSP70 PE 3 SV 1	1481.25	1 (0 ± 0.19) [0.44]
C6TGA6_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	227.22	1 (0 ± 0.93) [0.49]
Q9XET0_SOYBN	Putative uncharacterized protein OS Glycine max GN PM30 PE 2 SV 1	5145.05	1 (0 ± 0.15) [0.55]
C6T1V2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	727.8	1.01 (0.01 ± 0.44) [0.54]
C6TAA6_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	1169.44	1.02 (0.02 ± 0.63) [0.49]
C6SYU0_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	422.08	1.05 (0.05 ± 0.37) [0.64]
C6K8D1_SOYBN	Seed biotinylated protein 68 kDa isoform OS Glycine max PE 2 SV 1	782	1.05 (0.05 ± 0.19) [0.69]
B1Q2X4_SOYBN	Protein disulfide isomerase OS Glycine max GN PDIL 1 PE 3 SV 1	190.14	1.06 (0.06 ± 0.38) [0.58]
C6TNU2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	684.36	1.07 (0.07 ± 1.03) [0.6]
GLYG1_SOYBN	Glycinin G1 OS Glycine max GN GY1 PE 1 SV 2	34 638.61	1.07 (0.07 ± 0.07) [0.98]
Q6RIB7_SOYBN	Enolase OS Glycine max PE 2 SV 1	518.48	1.09 (0.09 ± 0.29) [0.68]
C6T588_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	7076.11	1.13 (0.12 ± 0.17) [0.92]
C6TCR6_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	163.95	1.13 (0.12 ± 0.47) [0.65]
C6T9Z5_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	11 699.31	1.15 (0.14 ± 0.12) [0.99]
LEC_SOYBN	Lectin OS Glycine max GN LE1 PE 1 SV 1	11 981.52	1.16 (0.15 ± 0.06) [1]
C6TB70_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	3485.42	1.16 (0.15 ± 0.19) [0.89]
C6TK76_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	1249.91	1.16 (0.15 ± 0.72) [0.62]
C6TKJ5_SOYBN	Triosephosphate isomerase OS Glycine max PE 2 SV 1	2573.78	1.19 (0.17 ± 0.26) [0.84]
C6TB67_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	1790.65	1.25 (0.22 ± 0.26) [0.92]
Q9ARI1_SOYBN	Lipoxygenase Fragment OS Glycine max PE 2 SV 1	5806.59	1.27 (0.24 ± 0.59) [0.76]
C6SXU0_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	295.31	1.27 (0.24 ± 0.38) [0.81]
Q7XAC5_SOYBN	Embryo-specific urease OS Glycine max PE 2 SV 1	128.19	1.27 (0.24 ± 0.49) [0.84]
Q9XER5_SOYBN	Seed maturation protein PM22 OS Glycine max GN PM22 PE 2 SV 1	7963.38	1.3 (0.26 ± 0.13) [1]
C6TNT2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	680.48	1.35 (0.3 ± 0.97) [0.78]
P93165_SOYBN	Em protein OS Glycine max PE 4 SV 1	4895.4	1.36 (0.31 ± 0.5) [0.91]
O22121_SOYBN	β-Subunit of β conglycinin Fragment OS Glycine max PE 2 SV 2	26 903.96	1.36 (0.31 ± 0.04) [1]
C6T8Q2_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	1063.38	1.39 (0.33 ± 1.17) [0.8]
O64458_SOYBN	Gly m Bd 30K allergen OS Glycine max GN P34 PE 2 SV 1	2154.71	1.46 (0.38 ± 0.2) [1]
C6EVF9_SOYBN	Elongation factor 1 α OS Glycine max GN EF 1A PE 2 SV 1	1653.6	1.49 (0.4 ± 0.16) [1]
C6TCF1_SOYBN	Putative uncharacterized protein OS Glycine max PE 2 SV 1	2055.68	1.62 (0.48 ± 0.75) [0.94]
LOX3_SOYBN	Seed lipoxygenase 3 OS Glycine max GN LOX1 3 PE 1 SV 1	21 758.1	1.75 (0.56 ± 0.12) [1]
Q96444_SOYBN	Actin Fragment OS Glycine max GN Soy119 PE 3 SV 1	662.59	1.75 (0.56 ± 1.42) [0.84]
Q39816_SOYBN	7S storage protein α subunit OS Glycine max PE 2 SV 1	11 664.26	1.82 (0.6 ± 0.18) [1]
Q50JD8_SOYBN	β-Conglycinin β subunit Fragment OS Glycine max PE 2 SV 1	26 292.45	1.84 (0.61 ± 0.11) [1]
SBP_SOYBN	Sucrose-binding protein OS Glycine max GN SBP PE 1 SV 1	4608.48	1.93 (0.66 ± 0.1) [1]
Q8GV24_SOYBN	Nucleoside diphosphate kinase OS Glycine max PE 2 SV 1	442.06	1.95 (0.67 ± 1.16) [0.88]
Q948X9_SOYBN	β-Conglycinin α subunit OS Glycine max PE 2 SV 1	18 794	2.08 (0.73 ± 0.06) [1]
Q96442_SOYBN	Actin Fragment OS Glycine max GN Soy57 PE 3 SV 1	1118.66	2.1 (0.74 ± 1.37) [0.88]
A7LCD6_SOYBN	Lipoxygenase OS Glycine max GN LOX10 PE 2 SV 1	419.03	2.12 (0.75 ± 0.39) [1]
Q7GC77_SOYBN	Glycinin A3B4 subunit OS Glycine max PE 1 SV 1	14 016.68	2.2 (0.79 ± 0.35) [1]
Q9SB11_SOYBN	Glycinin OS Glycine max GN A5A4B3 PE 2 SV 1	13 858.51	2.2 (0.79 ± 0.05) [1]
Q9SP11_SOYBN	Sucrose-binding protein homolog S 64 OS Glycine max GN SBP PE 2 SV 1	1005.86	2.29 (0.83 ± 0.44) [1]
GLYG4_SOYBN	Glycinin G4 OS Glycine max GN GY4 PE 1 SV 1	13 834.81	2.41 (0.88 ± 0.07) [1]
O04299_SOYBN	Elongation factor 1 α Fragment OS Glycine max GN TefS1 PE 4 SV 1	957.33	2.59 (0.95 ± 0.68) [1]
Q6V7V6_SOYBN	34 kDa maturing seed vacuolar thiol protease Fragment OS Glycine max PE 2 SV 1	1671.61	2.97 (1.09 ± 0.76) [0.97]
B3TDK9_SOYBN	Lipoxygenase OS Glycine max PE 3 SV 1	1434.12	3 (1.1 ± 0.2) [1]
GLYG2_SOYBN	Glycinin G2 OS Glycine max GN Gy2 PE 1 SV 2	26 155.1	3.46 (1.24 ± 0.05) [1]

Table 3. Continued

Accession	Description	Score	Soy Cn: Soy CTAG L3
HPSE_SOYBN	Hydrophobic seed protein OS Glycine max PE 1 SV 1	17 254.79	3.53 (1.26 + 0.63) [0.99]
7SB1_SOYBN	Basic 7S globulin OS Glycine max GN BG PE 1 SV 2	192.05	6.69 (1.9 + 0.29) [1]
Q43446_SOYBN	Lipoxygenase OS Glycine max GN vlxB PE 1 SV 1	2417.81	7.61 (2.03 + 0.52) [1]
Q39874_SOYBN	Lipoxygenase Fragment OS Glycine max PE 2 SV 1	10 427.04	12.81 (2.55 + 0.33) [1]
ACT1_SOYBN	Actin 1 OS Glycine max GN SAC1 PE 3 SV 2	1152.62	21.54 (3.07 + 0.89) [1]
GLYG5_SOYBN	Glycinin OS Glycine max PE 1 SV 1	8659.8	26.84 (3.29 + 0.23) [1]
Q39858_SOYBN	Soybean glycinin A3 B4 subunit Fragment OS Glycine max PE 2 SV 1	7988.31	30.27 (3.41 + 0.29) [1]
Q852U4_SOYBN	Glycinin A1bB2 784 OS Glycine max PE 2 SV 1	12 395.54	49.4 (3.9 + 0.41) [1]

reproducibility between different injections, usually in the full loop method, with 2 μL or 5 μL sample injection loading.

Figure 2 shows the resulting nanoUPLC chromatogram, the MS^E spectra from the $[M+2H]^{+2} = 857.87$ CTAG fragment and the respective spectra processed using PLGS. Five peptides from CTAG (Table 1, Fig. 1) were detected with high selectivity and specificity. These peptides showed no trace of post-translational modification; however, that possibility cannot be discarded because an additional 6 CTAG peptides were not detected. The orthogonal separations [48] obtained using SCX columns [49, 50] or recent technologies at the first-dimension linear gradient with fractions at different pH levels and with high-resolution separations in both the first and the second dimensions [51] were permitted because of the complexity of the chromatogram in this particular sample (Fig. 2A). Other methods for protein expression, including label-free with a continuous ion current and parallel fragmentation, have been used extensively for the monitoring of protein expression and have been described elsewhere [52].

3.2 Expression analysis

Additionally, a proteomic profile was processed with absolute quantitative values for each protein (Table 2). The CTAG recombinant protein was detected and quantified in nanograms based on the stoichiometric ion intensity values of the minimum three prototypic peptides of ADH and the identified protein. A relationship between the total detected protein and the specific protein concentration was determined, which allowed for the calculation of the percentage of the expressed protein in relation to TSP. The percentages of each detected protein are listed in Table 2. CTAG had an expression value of 0.1%, which is low compared with that of the other transgenic soybean seeds that express hGH [9] (2.9%), but it had a similar value compared with factor IX expression (0.2%) [10]. Other soybean proteins, such as β-conglycinin and glycinin, have been shown to exhibit the expected values that would be predicted for these storage proteins in soybean seeds [53].

Using this method, it was also possible to check protein expression regulation by comparing two or more samples. Table 3 shows a two-by-two comparison of the Soy CTAG L3 and Soy Cn protein expression list. This technique can also be used to check higher and lower regulations of native proteins, which are similar to the functions of the Identity^E and Expression^E software in PLGS (Waters) and provide information regarding the side effects at the proteomic level of the introduction of transgenes. In this case, a maturation protein, p24 oleosin, and a β conglycinin fragment were down-regulated in the Soy CTAG L3 transgenic line (Table 3). However, upregulations of glycinin, lectin and actin 1 were observed in the transgenic line (Table 3). This information extends the range of the technique and helps confirm that CTAG was expressed only in the Soy CTAG L3 sample.

4 Concluding remarks

This is an easy-to-follow method of determining whether a target recombinant protein is expressed in any expression system, especially in a situation where a small sample must be used or no antibody is available to perform blotting detection methods. Two samples, Soy CTAG L3 and Soy Cn-negative from the BR-16 cultivar, were used. The results indicate that Soy CTAG L3 was present in the sample and represented 0.1% of the TSP sample, which was low compared with the yields of the other recombinant proteins that have been expressed in soybean. The proteomic profile was also processed with absolute quantitation for each identified protein, which enabled the identification and analysis of higher and lower regulations of native proteins.

The authors are grateful to G. Ritter at the Ludwig Cancer Research Institute (New York Branch) for providing gene sequences and antibodies. They acknowledge support from C. Bloch at the Mass Spectrometry Laboratory-EMBRAPA. They acknowledge discussions with G. Ritter and C. Bloch and thank J. Taquita for technical help. This work was supported by the Brazilian Agricultural Research Corporation, the National Council for Scientific and Technological Development (CNPq) and the Fundação de Apoio a Pesquisa-DF (FAP-DF).

The authors have declared no conflict of interest.

5 References

[1] Swartz, J. R., *Curr. Opin. Biotechnol.* 2001, *12*, 195–201.

[2] Chu, L., Robinson, D. K., *Curr. Opin. Biotechnol.* 2001, *12*, 180–187.

[3] Tremblay, R., Wang, D., Jevnikar, A. M., Ma, S., *Biotechnol. Adv.* 2010, *28*, 214–221.

[4] Daniell, H., Singh, N. D., Mason, H., Streatfield, S. J., *Trends Plant Sci.* 2009, *14*, 669–679.

[5] Boothe, J., Nykiforuk, C., Shen, Y., Zaplachinski, S., Szarka, S., Kuhlman, P., Murray, E., Morck, D., Moloney, M. M., *Plant Biotechnol. J.* 2010, *8*, 588–606.

[6] Cunha, N. B. d., Araújo, A. C. G., Leite, A., Murad, A. M., Vianna, G. R., Rech, E. L., *Genet. Mol. Res.* 2010, *9*, 1163–1170.

[7] Jolliffe, N. A., Craddock, C. P., Frigerio, L., *Biochem. Soc. Trans.* 2005, *33*, 1016–1018.

[8] Ma, J. K.-C., Drake, P. M. W., Christou, P., *Nat. Rev. Genet.* 2003, *4*, 794–805.

[9] Cunha, N. B., Murad, A. M., Cipriano, T. M., Araujo, A. C. G., Aragao, F. J. L., Leite, A., Vianna, G. R., McPhee, T. R., Souza, G. H. M. F., Waters, M. J., Rech, E. L., *Transgenic Res.* 2011, *20*, 811–826.

[10] Cunha, N. B., Murad, A. M., Ramos, G. L., Maranhao, A. Q., Brỳgido, M. M., Araujo, A. C. G., Lacorte, C., Aragao, F. J. L., Covas, D. T., Fontes, A. M., Souza, G. H. M. F., Vianna, G. R., Rech, E. L., *Transgenic Res.* 2011, *20*, 841–855.

[11] Rech, E. L., Vianna, G. R., Aragão, F. J. L., *Nat. Protoc.* 2008, *3*, 410–418.

[12] Blas, A. L. D., Cherwinski, H. M., *Anal. Biochem.* 1983, *133*, 214–219.

[13] Perlmann, P., Engvall, E., *Immunochemistry* 1971, *8*, 871–874.

[14] O'Farrells, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[15] Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., Mann, M., *Nat. Protoc.* 2006, *1*, 2856–2860.

[16] Weiss, W., Görg, A., *Methods Mol. Biol.* 2007, *355*, 121–143.

[17] Blackstock, W. P., Weir, M. P., *Trends Biotechnol.* 1999, *17*, 121–127.

[18] Murad, A. M., Laumann, R. A., Lima, T. d. A., Sarmento, R. B. C., Noronha, E. F., Rocha, T. L., Valadares-Inglis, M. C., Franco, O. L., *Comp. Biochem. Physiol. C* 2006, *142*, 365–370.

[19] Murad, A. M., Noronha, E. F., Miller, R. N. G., Costa, F. T., Pereira, C. D., Mehta, Â., Caldas, R. A., Franco, O. L., *Microbiology* 2008, *154*, 3766–3774.

[20] Halligan, B. D., *Methods Mol. Biol.* 2009, *527*, 283–298.

[21] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

[22] Wilson, N., Simpson, R., Cooper-Liddell, C., *Methods Mol. Biol.* 2009, *534*, 205–212.

[23] Gevaert, K., Goethals, M., Martens, L., Damme, J. V., Staes, A., Thomas, G. R., Vandekerckhove, J., *Nat. Biotechnol.* 2003, *21*, 566–569.

[24] Hunter, A. P., Games, D. E., *Rapid Commun. Mass. Spectrom.* 1994, *8*, 559–570.

[25] Wilkins, J. A., Xiang, R., Horváth, C., *Anal. Chem.* 2002, *74*, 3933–3941.

[26] Guerrier, L., Righetti, P. G., Boschetti, E., *Nat. Protoc.* 2008, *3*, 883–890.

[27] Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., Smith, R. D., *Anal. Chem.* 2002, *74*, 4235–4249.

[28] Nouri, M.-Z., Komatsu, S., *Proteomics* 2010, *10*, 1930–1945.

[29] Brumbarova, T., Matros, A., Mock, H.-P., Bauer, P., *Plant J.* 2008, *54*, 321–334.

[30] Behrens, H. L., Chen, R., Li, L., *Anal. Chem.* 2008, *80*, 6949–6958.

[31] Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., Wijk, K. J. v., *PLoS One* 2008, *3*, e1994.

[32] Barnes, S., Shonsey, E. M., Eliuk, S. M., Stella, D., Barrett, K., Srivastava, O. P., Kim, H., Renfrow, M. B., *Biochem. Soc. Trans.* 2008, *36*, 1037–1044.

[33] Bouché, J.-P., Froment, C., Dozier, C., Esmenjaud-Mailhat, C., Lemaire, M., Monsarrat, B., Burlet-Schiltz, O., Ducommun, B., *J. Proteome Res.* 2008, *7*, 1264–1273.

[34] Unwin, R. D., Griffiths, J. R., Whetton, A. D., *Nat. Protoc.* 2009, *4*, 870–877.

[35] Mori, M., Abe, K., Yamaguchi, H., Goto, J., Shimada, M., Mano, N., *J. Proteome Res.* 2010, *9*, 3741–3749.

[36] Yang, Y., Zhang, S., Howe, K., Wilson, D. B., Moser, F., Irwin, D., Thannhauser, T. W., *J. Biomol. Tech.* 2007, *18*, 226–237.

[37] Levin, Y., Wang, L., Ingudomnukul, E., Schwarz, E., Baron-Cohen, S., Palotás, A., Bahn, S., *J. Chromatogr. B* 2009, *877*, 1299–1305.

[38] Li, G.-Z., Vissers, J. P. C., Silva, J. C., Golick, D., Gorenstein, M. V., Geromanos, S. J., *Proteomics* 2009, *9*, 1696–1719.

[39] Geromanos, S. J., Vissers, J. P. C., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V., Bateman, R. H., Langridge, J. I., *Proteomics* 2009, *9*, 1683–1695.

[40] Xu, D., Suenaga, N., Edelmann, M. J., Fridman, R., Muschel, R. J., Kessler, B. M., *Mol. Cell. Proteomics* 2008, *7*, 2215–2228.

[41] Cheng, F.-y., Blackburn, K., Lin, Y.-m., Goshe, M. B., Williamson, J. D., *J. Proteome Res.* 2009, *8*, 82–93.

[42] Gnjatic, S., Nishikawa, H., Jungbluth, A. A., Güre, A. O., Ritter, G., Jäger, E., Knuth, A., Chen, Y.-T., Old, L. J., *Adv. Cancer Res.* 2006, *95*, 1–30.

[43] Chen, Y., Scanlan, M. J., Sahin, U., Türeci, Ö., Gure, A. O., Tsang, S., Williamson, B., Stockert, E., Pfreundschuh, M., Old, L. J., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 1914–1918.

[44] Kurashige, T., Noguchi, Y., Saika, T., Ono, T., Nagata, Y., Jungbluth, A., Ritter, G., Chen, Y.-T., Stockert, E., Tsushima, T., Kumon, H., Old, L. J., Nakayama, E., *Cancer Res.* 2001, *61*, 4671–4674.

[45] Murphy, R., Green, S., Ritter, G., Cohen, L., Ryan, D., Woods, W., Rubira, M., Cebon, J., Davis, I. D.,

Sjolander, A., Kypridis, A., Kalnins, H., McNamara, M., Moloney, M. B., Ackland, J., Cartwright, G., Rood, J., Dumsday, G., Healey, K., Maher, D., Maraskovsky, E., Chen, Y., Hoffman, E. W., Old, L. J., Scott, A. M., *Prep. Biochem. Biotechnol.* 2005, *35*, 119–134.

[46] Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G.-Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., Geromanos, S., *Anal. Chem.* 2005, *77*, 2187–2200.

[47] Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., Geromanos, S. J., *Mol. Cell. Proteomics* 2006, *5*, 144–156.

[48] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., *Anal. Chem.* 2005, *77*, 6426–6434.

[49] Millea, K. M., Kass, I. J., Cohen, S. A., Krull, I. S., Gebler, J. C., Bergerb, S. J., *J. Chromatogr. A* 2005, *1079*, 287–298.

[50] Gilar, M., Olivova, P., Chakraborty, A. B., Jaworski, A., Geromanos, S. J., Gebler, J. C., *Electrophoresis* 2009, *30*, 1157–1167.

[51] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., *J. Sep. Sci.* 2005, *28*, 1694–1703.

[52] Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Murrell, A. C. J. C., Scrivens, J. H., *J. Proteome Res.* 2009, *8*, 3752–3759.

[53] Li, C., Zhang, Y.-M., *Heredity* 2010, in press. DOI: 10.1038/hdy.2010.97.

