Does filter aided sample preparation (FASP) provide sufficient method linearity for quantitative plant shotgun proteomics?

Tatiana Leonova,[1,2] Christian Ihling,[3] Mohamad Saoud,[1] Robert Rennert,[1] Ludger A. Wessjohann,[1] Andrej Frolov [1,2*]

[1]Leibniz Institute of Plant Biochemistry, Department of Bioorganic Chemistry, [2]St. Petersburg State University, Department of Biochemistry, [3]Martin-Luther Universität Halle-Wittenberg, Institute of Pharmacy, Department of Pharmaceutical Chemistry and Bioanalytics

*Corresponding author:

Dr. Andrej Frolov

Leibniz Institute of Plant Biochemistry

Department of Bioorganic Chemistry

Weinberg 3

06120 Halle (Saale), Germany

Tel. +49(345) 55821350

Fax. +49(345) 55821309

Email: afrolov@ipb-halle.de

Abstract

Gel-free LC-based shotgun proteomics represents the current gold standard of proteome analysis due to its outstanding throughput, analytical resolution and reproducibility. Thereby, the efficiency of sample preparation, i.e., protein isolation, solubilization and proteolysis, directly affects the correctness and reliability of quantification, being therefore the bottle neck of shotgun proteomics. The desired performance of the sample preparation protocols can be achieved by application of detergents. However, these ultimately compromise reverse phase chromatographic separation and disrupt electrospray ionization. Filter aided sample preparation (FASP) represents an elegant approach to overcome these limitations. Although this method is comprehensively validated for cell proteomics, its applicability to plants and compatibility with plant-specific protein isolation protocols is still unknown, i.e., no data on linearity of underlying protein quantification methods for plant matrices is available. To fill this gap, we address here the potential of FASP in combination with two protein isolation protocols for quantitative analysis of pea (*Pisum sativum*) seed and *Arabidopsis thaliana* leaf proteomes by the shotgun approach. For this, in comprehensive spiking experiments with bovine serum albumin (BSA), we evaluated the linear dynamic range (LDR) of protein quantification in the presence of plant matrices. Further, we addressed the interference of two different plant matrices in quantitative experiments, accomplished with two alternative sample preparation workflows in comparison to conventional FASP-based digestion of cell lysates, considered here as a reference. Our results indicate very good applicability of FASP to quantitative plant proteomics with an only limited impact of the protein isolation technique on the methods overall performance.

Keywords: Detergent-assisted proteolysis, filter aided sample preparation (FASP), label-free quantification, LC-MS, phenol extraction, plant proteomics, shotgun proteomics, sodium dodecyl sulfate

1 Introduction

To date, bottom-up proteomics represents one of the most established methodological platforms for post-genomic research.[1] During the last decade, gel-free LC-based shotgun proteomics became a gold standard of proteome analysis due to its higher throughput, superior proteome coverage, better analytical resolution and reproducibility.[2] However, because of the mechanistic limitations of electrospray ionization (ESI), shotgun proteomics is critically sensitive to detergents.[3] On the other hand, in contrast to in-gel proteolysis (which can be quantitatively accomplished in ammonium bicarbonate buffer), in-solution digestion for shotgun proteomics ultimately requires supplementation of detergents to ensure quantitative solubilization of protein isolates and their efficient proteolysis.[4] Therefore, a broad range of protocols for detergent-assisted proteolysis, employing degradable or selectively removable detergents were successfully established to date.[5] One of the most widely spread one is the filter aided sample preparation (FASP) approach – an elegant method efficiently combining the advantages of in-gel and in-solution digestion protocols.[4]

This method allows complete solubilization of dried protein isolates in sodium dodecyl sulfate (SDS) aqueous solutions, centrifugal concentration of reconstituted proteins, efficient reduction and alkylation followed with detergent removal and digestion of proteins in one centrifugal filter device. After introduction of FASP by the Mann group in 2009,[4] the original protocol was subjected to various modifications to improve recovery of proteolytic peptides.[6] Thus, binding of peptides to the membrane could be reduced by conditioning of filters with 5% (*v/v*) Tween® 20,

whereas supplementation of 0.2% (*w/v*) deoxycholic acid prior to proteolysis resulted in enhancement of digestion efficiency.[7] Further, implementation of a multienzyme digestion FASP provided improved protein identification rates and sequence coverage of individual species.[8]

The principal advantage of FASP technology is its ability to couple uniform and efficient protein extraction and/or reconstitution protocols with powerful proteolysis techniques. In this context, due to the tremendous variation in the properties of biological matrices from different (plant) species, the entire sample preparation protocol typically requires intensive optimization for each of them to ensure the best possible performance.[9] For this reason, integration of FASP with protein extraction techniques still requires validation, which typically relies on the assessment of proteome coverage in comparison to other sample preparation pipelines.[6] In these experiments, FASP (which was originally proposed as the method for digestion of cell lysates[4]) was shown to be widely applicable to human and animal samples.[8] Moreover, this method was successfully employed in proteomics analyses of plant organs – leaves,[10] seeds,[11] roots,[12] and fruits.[13] Thereby, FASP proved to be compatible with all three major protein isolation strategies – phenol extraction, precipitation with TCA/acetone, and their combination.[10,14]

Although the efficiency of FASP for plant samples was characterized in terms of protein identification rates and sequence coverage, the potential of this technique for quantitative plant proteomics remains completely unknown. Indeed, high contents of proteases, carbohydrates and secondary metabolites, including protein-binding polyphenols, characteristic for recalcitrant plant tissues, might interfere with protein extraction and MS analysis, dramatically affecting, thereby, the linear dynamic range (LDR) of protein quantification. However, the information about LDR of protein quantification by FASP-based bottom-up shotgun proteomics is still missing. Moreover, the impact of specific protein isolation techniques in the overall result of

such experiments is also unknown. Hence, the FASP-based sample preparation methods still require validation in terms of their applicability for quantitative assessments. Therefore, to fill this gap, we addressed the potential of FASP in combination with two protein isolation protocols for quantification of *Arabidopsis thaliana* leaf and pea (*Pisum sativum* L.) seed proteins by LC-MS-based bottom-up shotgun proteomics. We evaluated the LDR of protein quantification in the presence of complex plant matrices processed by two alternative sample preparation workflows, each in comparison to conventional FASP-based digestion protocols of cell lysates. Our results indicate the applicability of FASP for quantitative plant proteomics with a limited impact of protein isolation technique used on the overall method performance.

2 Materials and methods

2.1 Materials and reagents

Materials were obtained from the following manufacturers: Biowest (Nuaillé, France): Fetal Bovine Serum, South America; Capricorn Scientific GmbH (Ebsdorfergrund, Germany): RPMI 1640, Dulbecco's PBS (1x), Penicillin/Streptomycin (100x), L-Glutamine solution (200 mmol/L), Trypsin-EDTA (0.05 %) in DPBS (1x); Carl Roth GmbH and Co (Karlsruhe, Germany): tris(hydroxymethyl)aminomethane (tris, ultra-pure grade), tetramethylethylenediamine (TMED, p.a.), ammonium persulfate (ACS grade), glycerol (p.a.), bovine serum albumin (BSA); CDS Analytical, LLC (Oxford, PA, USA): Empore Extraction C18 Disks; Honeywell (Charlotte, NC, USA): acetonitrile (LC-MS grade), methanol (LC-MS grade); PanReac AppliChem (Darmstadt, Germany): glycerol (ACS grade), phenylmethylsulfonyl fluoride (PMSF), polysorbate 20 (Tween® 20); SERVA Electrophoresis GmbH (Heidelberg, Germany): 2-mercaptoethanol (research grade), acrylamide/bis-acrylamide solution (37.5/1, 30% (*w/v*), 2.6% C), NB sequencing grade modified trypsin from porcine

pancreas, sodium dodecyl sulfate (SDS, electrophoresis grade); Thermo Fisher Scientific (Bremen, Germany): PageRuler™ Prestained Protein Ladder #26616 (10–180 kDa); VWR Chemicals, LLC (Solon, OH, USA): phenol (ultra-pure). Amicon® Ultra-0.5 Centrifugal Filter Unit of 30 kDa molecular weight cutoff (MWCO) and all other chemicals were purchased from Merck KGaA (Darmstadt, Germany).

The PC-3 cell line was maintained routinely in complete RPMI medium 1640 supplemented with 10% of heat-inactivated FBS, 1% glutamine and 1% penicillin/streptomycin at 37 °C in a humidified atmosphere with 5% CO_2 to reach subconfluency (~ 80%) prior to subsequent usage or subculturing.[15] Pea seeds of the cultivar "Millennium" were obtained from the Research and Practical Center of the National Academy of Science of the Republic of Belarus for Arable Farming (Zhodino, Belarus, harvested in the year 2015 and stored at 18ºC). *Arabidopsis thaliana* (Columbia 1092) seeds were planted in wet soil-sand mixture, and the plants were grown in a phytotron MLR-351H (Sanyo Electric Co., Ltd., Moriguchi, Japan) under a short day with a 8 h light (150 ± 2.5 µmol photons m^{-2} s^{-1})/16 h dark cycle at 23/18 °C, respectively, and 60% humidity. The plants were harvested, and the leaves were ground in liquid nitrogen using a Mixer Mill MM 400 ball mill with 3-mm-diameter stainless steel balls (Retsch, Haan, Germany) at a vibration frequency of 30 Hz for 2 min. The plants were characterized elsewhere.[16]

2.2 Protein isolation

Protein isolation relied on two approaches: (*i*) the phenol extraction method described in detail previously[17] and (*ii*) treatment with an SDS-based solution according to the procedure recently described by Bassal et al[18] with minor modifications. In detail, protein isolation from PC-3 cells and a mixture of pea seed powder and Arabidopsis leaf material was accomplished by treatment with extraction buffer (4% (*w/v*) SDS, 10 mmol/L dithiothreitol (DTT), 10 mmol/L EDTA, 1

mmol/L PMSF in 50 mmol/L Tris-HCl, pH 8.0), followed by two steps of incubation (900 rpm; at 95°C for 10 min, at room temperature for 20 min) and centrifugation (25 000 g, 30 min, 10°C) after each incubation. The resulting supernatants were transferred to new tubes.

The total protein fractions from pea seeds and mixtures of pea seed powder and Arabidopsis leaf material were isolated by the phenol extraction procedure described in detail previously,[17] and dry protein pellets were reconstituted in 10% (*w/v*) SDS solution. The protein contents were determined by the 2D Quant Kit according to the manufacturer instructions.

2.3 FASP protocol

The Amicon® Ultra 30K filter units were conditioned (passivated) on the day prior to digestion with 5% (*v/v*) aqueous Tween® 20 under continuous shaking overnight. Afterwards, filters were washed three times for 10 min with distilled water. The protein aliquots (50 µg) were adjusted to the total volume of 200 µL with urea solution (8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0), applied to the filter unit and centrifuged (here and below – 14 000 g, 10 min). The concentrated samples were washed three times with 200 µL of urea solution followed each time by centrifugation. Disulfide bonds were reduced by addition of 100 µL of a solution of 100 mmol/L DTT, 8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0 and incubation at 22°C for 1 h under continuous shaking (450 rpm) followed by centrifugation. Alkylation of sulfhydryl groups was accomplished by addition of 100 µL of 50 mmol/L iodoacetamide in a solution of 8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0 and incubation at 22°C for 1 h in the dark under continuous shaking followed by centrifugation. The resulted concentrated samples were washed three times with 200 µL of urea solution and three times with 100 µL of 50 mmol/L aq. NH_4HCO_3 followed each time by centrifugation. Afterwards, the proteins were digested by sequential addition of two aliquots of 2.5 and 1 µg trypsin, reconstituted in 50 µL of 50 mmol/L aq. NH_4HCO_3 (enzyme to

protein ratio 1:20 and 1:50 *w/w,* respectively) and incubation at 37°C for 4 and 12 h under continuous shaking (450 rpm), respectively. The digests were collected by centrifugation, and the filter units were rinsed with 40 µL of 50 mmol/L aq. NH_4HCO_3 three times followed by centrifugation. The resulting filtrates were desalted by solid phase extraction (SPE) as described by Mamontova et al.[19] The completeness of tryptic digestion was verified by SDS-PAGE as described by Greifenhagen et al.[20]

2.6 LC-MS/MS analysis

All samples were analyzed by nanoRP-HPLC-ESI-MS/MS using an Orbitrap XL hybrid mass spectrometer equipped with a NanoFlex source, coupled online to an Ultimate 3000 nano-HPLC system (Thermo Fisher Scientific, Bremen, Germany). Proteolytic peptides were loaded on a trap column (PepMap 100 C18, 300 µm × 5 mm, particle size 5 µm) during 15 min with 0.1 % (*v/v*) TFA at a flow rate of 30 µL/min and resolved on a separation column (PepMap 100 C18, 75 µm × 150 mm, particle size 3 µm, Thermo Fisher Scientific, Bremen, Germany). The peptides were separated with linear gradient 3–35% eluent B over 90 min (A: 0.1% *v/v* aqueous formic acid, B: 0.08% *v/v* formic acid in acetonitrile) at the flow rate of 300 nL/min. The raw files were acquired as data-dependent acquisition (DDA) experiments accomplished in positive ion mode. Dependent tandem mass spectrometric (MS/MS) experiments relied on higher energy collision-induced dissociation (HCD) at 27% normalized collision energy (NCE). MS data (*m/z* range 300-1500) were recorded with R = 120 000, the target of the automated gain control (AGC) was set to 2 x 10^5 and the maximum injection time was 50 ms. Each full scan was followed by high-resolution HCD product ion scans within 5 s, starting with the most intense signal in the mass spectrum, with charge states ranging from 2 to 6. For MS/MS scans, the following parameters were applied: resolution (R) of 15 000, AGC of 5 × 10^4 and maximum injection time of 200 ms.

Dynamic exclusion of multi-charged peptide ions was set to 60 s. The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE[21] partner repository with the dataset identifier PXD025897 and 10.6019/PXD025897.

2.7 Data analysis

Identification of peptides relied on Proteome Discoverer software (version 2.2.0.388, Thermo Fisher Scientific, Bremen, Germany) and Sequest HT search engine. For database search, the enzyme was set to trypsin, tolerating two missed cleavages. The precursor and fragment mass tolerance were set to 10 ppm and 0.8 Da, respectively. Carbamidomethylation of cysteine was employed as a fixed modification, and oxidation of methionine was specified as a variable one. False discovery rate (FDR) was set to 0.05. Proteotypic peptides were selected for integration if they were confidently (XCorr $\geq$ 2.20) annotated as $[M+2H]^{2+}$ ions, did not contain missed cleavage sites, modifications and methionine or cysteine amino acid residues. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific). Detection and integration of the peptide-specific peaks in corresponding extracted ion chromatograms (XICs, m/z $\pm$ 0.02) were accomplished by ICIS algorithm with following parameters: Smoothing Points 15, Baseline Window 40-80, Area Noise Factor 5, Peak Noise Factor 10. The linearity of protein quantification was assessed for several proteotypic peptides by Xcalibur software (version 2.0.7, Thermo Fisher Scientific, Bremen, Germany).

3 Results and discussion

3.1 Quantitative analysis of BSA spiked in cell lysates

Originally, FASP was proposed for (human/animal) cell lysates,[4] and its performance was comprehensively characterized with various cell lines and experimental setups.[4,22,23] Therefore, here we decided for cultured cells as a reference to estimate the linear performance of this

method. Following the previously established methodology, prostate cancer (PC-3) cells were lyzed with extraction buffer containing 4% (*w/v*) SDS, incubated at 95°C and centrifuged, i.e. the procedure reproduced the classical workflow of Wiśniewski et al.[4] Of course, the total cell lysate represents a complex system, highly prone to matrix effects (most probably – predominantly ion suppression),[24] which can affect the linear dynamic ranges (LDRs) of individual tryptic peptides.

To get access to the LDRs in the easiest and most straightforward way, we applied a spiking approach, which represents a well-established normalization strategy in label-free quantification (LFQ) experiments.[25] Specifically, after determination of protein concentrations, aliquots of cell lysates were spiked with BSA at the percentage concentration ratios of 3.125, 6.25, 12.5, 25, 50, 100% (*w/w*). This range of ratios covered two orders of magnitude, that essentially exceeded the magnitude of typical protein expression responses both in cell lysates and plant protein extracts,[26] and therefore is in agreement with our aims. Although spiking with mixtures of standard peptides represents an adequate approach to estimate method linearity performance,[27] we decided here for the spike with allogenic protein, as this approach might consider not only matrix effects, but also the contribution of factors related to proteolysis. Bovine serum albumin seems to be a suitable protein for this purpose. Indeed, it has 58 lysyl and 24 arginyl residues, which are relatively homogenously distributed in the protein sequence. It gives a rich selection of potential proteotypic peptides covering the whole range of peptide-specific retention times. Indeed, serum albumin was established as a tool for reference protein normalization both in cell[28] and plant[29] proteomics. Thus, the behavior of its proteolytic peptides in plant protein hydrolysates is well characterized and is in agreement with dynamic ranges of cell- and plant-derived tryptic peptides. This, in turn will ensure comparability of the results obtained with cell lysates and plant protein extracts. In agreement with earlier reports,[22,23] FASP proved to be an

efficient tool for the analysis of cell lysates. The completeness of tryptic digestion was verified by SDS-PAGE and also was in agreement with the results of previous studies.[4]

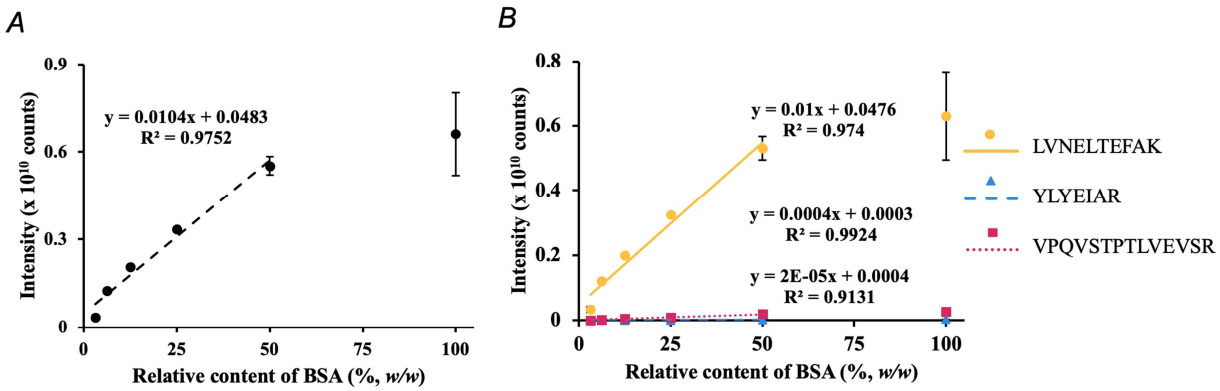

Figure 1 Assessment of methods linearity for quantification of bovine serum albumin (BSA) spiked to the prostate cancer (PC-3) cell lysate to obtain the BSA relative contents of 3.125, 6.25, 12.5, 25, 50, 100 % (*w/w*). Quantification of BSA relied on the sum (A) of integrated peak areas obtained for *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02 and 756.4250 ± 0.02 at t_R 52.4, 45.6 and 50.2, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides LVNELTEFAK (circles, solid line), YLYEIAR (triangles, dashed line) and VPQVSTPTLVEVSR (squares, dotted line), respectively. Quantification of individual peptides is shown on panel B. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

After the nanoRP-HPLC-ESI-LIT-Orbitrap-MS analysis, the raw files were searched by SEQUEST engine against the BSA sequence database. Sequence coverage of BSA spiked to cell lysates corresponded to 84% with 68 identified unique peptides. Based on this search, three proteotypic BSA peptides, namely LVNELTEFAK, AEFVEVTK and YLYEIAR with *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02 and 756.4250 ± 0.02 at t_R 78.1, 66.0 and 75.7, corresponding to the $[M+2H]^{2+}$ ions, respectively, were used to assess the linearity of the sample preparation

method. The method delivered acceptable linear correlation (R^2 = 0.975) over half of the assessed dynamic range (Figure 1A) based on the relative intensity of the overall BSA abundance response (calculated as an integrated sum of corresponding peptide signal intensities) and its contents spiked to aliquots of cell lysates. Accordingly, each of the individual peptides demonstrated an excellent linearity of response R^2 from 0.91 to 0.97 (Figure 1B).

Thus, in our hands, the FASP technique delivered a good performance with cell lysates, and our results were comparable with the classical works published before.[6] Based on this, we assume that the PC-3 cell lysate spiked with BSA is an appropriate reference for our plant shotgun proteomics experiments.

3.2 Quantitative analysis of BSA spiked to the protein extracts of the pea seeds

Having the verified FASP method in hand, we extended our spike approach to pea seed total protein fraction. For this, proteins were isolated from the plant tissues by phenol extraction, dry protein pellets were reconstituted in 10% (*w/v*) SDS solution and protein concentrations were determined. Based on these values, a 1 g/L BSA solution in 10% SDS was spiked to protein solution aliquots using the scheme described above for the cell lysates. Thus the digestion protocol of the Mann group was transferred to the appropriate protein extraction method, which is currently considered as the most efficient one in terms of sample quality and protein identification rates.[2]

The FASP approach proved to be an appropriate method for digestion of the pea seed proteins, i.e., it was ideally compatible with the phenol extraction method. Indeed, most of the total seed protein fraction (97.7%) was successfully digested and transferred through the cellulose membrane of the filter units (Figure 2A), whereas only 2.3% of the protein aliquot, subjected to proteolysis, remained on the membrane after 3x washing with 50 mmol/L aq. NH_4HCO_3 (as the

averaged total lane abundance of the non-filtered fraction constituted 23% of the ND reference, which, in turn, corresponded to 10% of the digested aliquot, Figure 2B).

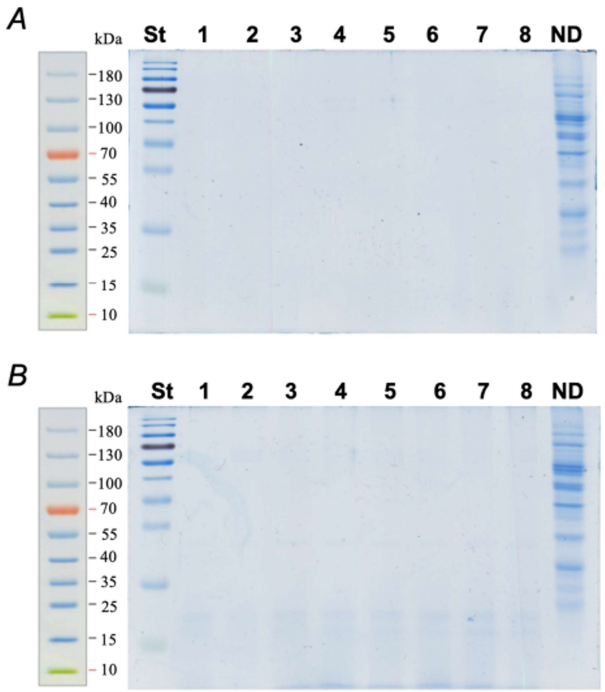

Figure 2 SDS-PAGE electropherograms acquired for the pea seed protein digested with trypsin using the FASP approach (in total 50 µg of protein were applied to each filter unit). The analyses were performed with filtrate (A) and the fraction retained on the filter (B) after 3 x washing with 40 µL of 50 mmol/L aq. NH_4HCO_3 and centrifugal filtration (14 000 g, 10 min). To assess the completeness of hydrolysis, 5 µg of each digest were applied on the gel. The whole retained fraction (corresponding to 50 µg of digested protein) was completely transferred to a polypropylene tube, lyophilized, reconstituted in SDS-PAGE sample buffer and loaded on the gel. The overall lane densities were compared to those of non-digested (ND) protein (5 µg) applied to a separate lane. St – Page Ruler Prestained Protein Ladder.

To address the efficiency of protein quantification with the FASP approach, the integrated peak areas were calculated for each peptide signal in corresponding extracted ion chromatograms

(XICs, m/z ± 0.02) and summed for each concentration point. This approach yielded the same LDR as was observed above for the PC-3 cells (Figure 3A). For all three selected proteotypic BSA peptides, similar concentration-signal intensity curves were observed (Figure 3B). Thereby, the response was linear for up to 50% of spiked BSA (R^2 = 0.97 - 0.99).

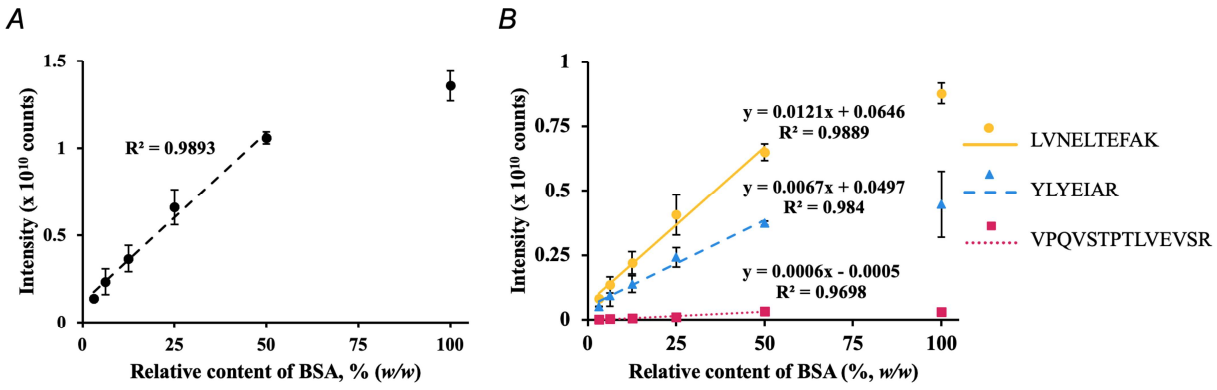

Figure 3 Assessment of method linearity for the quantification of bovine serum albumin (BSA) spiked to the total pea seed protein to obtain the BSA relative contents of 3.125, 6.25, 12.5, 25, 50, 100 % (*w/w*). Seed proteins were isolated by phenol extraction. Quantification of BSA relied on the sum (A) of integrated peak areas obtained for *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02 and 756.4250 ± 0.02 at t_R 78.1, 66.0 and 75.7, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides LVNELTEFAK (circles, solid line), YLYEIAR (triangles, dashed line) and VPQVSTPTLVEVSR (squares, dotted line), respectively. Quantification of individual peptides is shown on panel B. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

This result highlights the general applicability of FASP in quantitative plant proteomics, i.e. its combination with phenol extraction gave access to acceptable linearity for quantification of BSA spiked to plant protein isolates (Figure 3). Moreover, BSA spiked to pea seed protein yielded superior sequence coverage (92%) and a higher number of identified unique peptides (93) in

comparison to BSA spiked to PC-3 cell lysate. This fact can be explained by the higher efficiency of phenol extraction with respect to discrimination from non-protein contaminations.[30] On the other hand, this observation can be attributed to different relative abundances of individual proteins in these matrices. Indeed, pea seeds contain several strongly dominant proteins, whereas most proteins are much less abundant.[31] This might result in lower ion suppression at most of the chromatogram span.

3.3 Quantitative analysis of Arabidopsis proteins in pea seed protein matrix

In a next step we addressed the impact of the protein isolation method on the performance of FASP-based quantitative proteome analysis. To get access to this information, we compared the linearity of two protein isolation methods – (*i*) phenol extraction and (*ii*) treatment with a detergent-containing solution followed by incubation at 95°C. For this, frozen milled pea seed and Arabidopsis leaf material were co-extracted in different weight ratios (9:1, 3:1, 1:1, 1:3, 1:9, 0:1) using the phenol extraction procedure or treatment with SDS-containing extraction solution. The completeness of tryptic digestion was verified by SDS-PAGE (Supplementary information, Figure S1).

As protein yields from pea seeds were 8-fold higher in comparison to Arabidopsis leaf (5.4 vs 42.2 mg/g fresh weight), even the most abundant Arabidopsis proteins acted as minor components of the mixed protein isolates. Therefore, we followed the abundance of two proteins, characteristic for Arabidopsis leaves – large subunit of chloroplastic ribulose bisphosphate carboxylase/oxygenase (RuBisCO, the most abundant leaf polypeptide, Figure 4A) and RuBisCO activase, which is less abundant (Figure 4B). The first protein was quantified with the *m/z* 511.2693 ± 0.02, 614.8302 ± 0.02 and 704.3376 ± 0.02 at t_R 60.0, 49.4 and 44.6 corresponding to $[M+H]^{2+}$ ions of proteotypic peptides DTDILAAFR, DLAVEGNEIIR and

LTYYTPEYETK, respectively, whereas the second one – with *m/z* 504.2741 ± 0.02, 849.3843 ± 0.02 and 576.8606 ± 0.02 at t_R 42.2, 40.9 and 67.9, corresponded to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides FVESLGVEK, GLAYDTSDDQQDITR and VPLILGIWGGK, respectively.

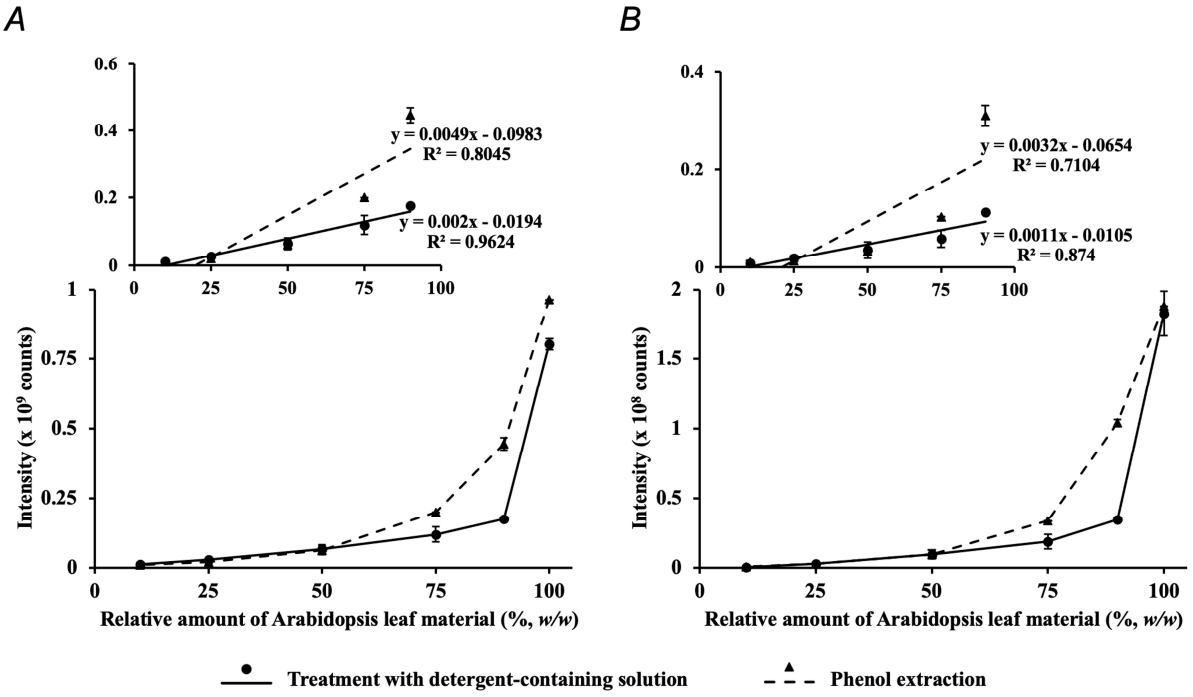

Figure 4 Assessment of the method linearity for quantification of *Arabidopsis thaliana* leaf proteins ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) and RuBisCO-activase after co-extraction from milled frozen pea seed and Arabidopsis leaf material in different weight ratios (9:1, 3:1, 1:1, 1:3, 1:9, 0:1) using the phenol extraction procedure (triangles, dashed line) or treatment with SDS-containing extraction solution (circles, solid line). Quantification of RuBisCO (A) relied on the sum of integrated peak areas obtained for *m/z* 511.2693 ± 0.02, 614.8302 ± 0.02 and 704.3376 ± 0.02 at t_R 60.0, 49.4 and 44.6, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides DTDILAAFR, DLAVEGNEIIR and LTYYTPEYETK, respectively, whereas RuBisCO activase (B) was quantified with *m/z* 504.2741 ± 0.02, 849.3843

± 0.02 and 576.8606 ± 0.02 at t_R 42.2, 40.9 and 67.9, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides FVESLGVEK, GLAYDTSDDQQDITR and VPLILGIWGGK, respectively. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

As can be seen from Figure 4, direct extraction with SDS-containing solution provided better linearity of the integrated response of Arabidopsis proteins, each based on a sum of the selected three proteotypic peptides. For all of them, this effect was determined with a compromised (in comparison to phenol extraction) recovery of the corresponding proteins in SDS-containing extraction buffer when Arabidopsis leaf material contributed more than 50% in the plant material mix (Supplementary information, Figure S2). Indeed, when the detergent solution was used for extraction, linearity was superior (R^2 values better than 0.95) up to a contribution of Arabidopsis material accounting 90%, whereas for phenol extraction R^2 values typically were 0.85 or lower. However, when the quantification dynamics range was reduced to a contribution of Arabidopsis material accounting 50%, both detergent and phenol extraction methods yielded similar linearity (R^2 values of 0.99 and 0.97, respectively, data not shown). Most probably, the observed differences are attributed to stronger matrix effects, which might accompany detergent-based extraction. Indeed, secondary metabolites, which are co-extracted with proteins in this design, cannot be quantitatively removed solely by ultrafiltration and might cause inhibition of trypsin activity and ion suppression.[10] On the other hand, phenol isolates are free from non-protein contaminants, and pea proteolytic peptides represented the only factor of ion suppression in our experimental system. Accordingly, the signal intensity of Arabidopsis tryptic peptides increased, when the contribution of pea seed protein in the total isolate decreased.

Thus, protein isolation protocol affects the linear dynamic range of FASP-based quantitative proteomics techniques. However, this difference in linear dynamic ranges can be considered during data interpretation and appropriately corrected by experimental design.

4 Conclusions

Filter aided sample preparation (FASP) represents a powerful and versatile technique to access quantitative and reproducible protein solubilization and digestion for shotgun proteomics. It was originally proposed for cell lysates in mid 2000s by Mann group. Since that time, it was comprehensively optimized and validated for cells, blood plasma and homogenates of animal organs. Finally, during recent years FASP was employed in plant proteomics as well. However, in our opinion this step requires a comprehensive estimation of the methods behavior with plant matrix, which is known to be much more complex and difficult in handling compared to mammalian cells. The most critical aspect here is the effect of the plant matrix on LDRs of individual proteins. This knowledge is critically important for a correct assessment of quantitative alterations. Here, to the best of our knowledge, for the first time, we provide data on the linearity of FASP in plant matrix. Surprisingly, when coupled to plant-specific protein isolation protocols, this method demonstrates even better performance in comparison to mammalian matrices. The selection of the protein isolation protocol for plant FASP assumes a compromise between recovery (which is more favored by the phenol extraction method) and LDR (which is better when direct detergent treatment is applied). Therefore, a linearity/recovery test prior to working with a new plant matrix is mandatory for obtaining adequate quantitative information.

5 Acknowledgements

We thank Dr. Wolfgang Höhenwarter for fruitful discussions. The financial support from the Russian Foundation of Basic Research (# 20-54-00044) and Russian Science Foundation (# 20-16-00086) is gratefully acknowledged.

Associated content

Supplementary file:

Figure S1 SDS-PAGE electropherograms acquired for the co-extracted Arabidopsis leaf and pea seed proteins digested with trypsin using the FASP approach (in total 50 µg of protein were applied to each filter unit).

Figure S2 Assessment of the method linearity for quantification of Arabidopsis leaf proteins ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO, A-C) and RuBisCO activase (D-F) after co-extraction from frozen milled pea seed and Arabidopsis leaf material in different weight ratios (9:1, 3:1, 1:1, 1:3, 1:9, 0:1) using the phenol extraction procedure or treatment with SDS-containing extraction solution.

References

(1) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews*. NIH Public Access April 10, 2013, pp 2343–2394. https://doi.org/10.1021/cr3003533.

(2) Smolikova, G.; Gorbach, D.; Lukasheva, E.; Mavropolo-Stolyarenko, G.; Bilova, T.; Soboleva, A.; Tsarev, A.; Romanovskaya, E.; Podolskaya, E.; Zhukov, V.; Tikhonovich, I.; Medvedev, S.; Hoehenwarter, W.; Frolov, A. Bringing New Methods to the Seed Proteomics Platform: Challenges and Perspectives. *International Journal of Molecular*

Sciences. MDPI AG December 1, 2020, pp 1–54. https://doi.org/10.3390/ijms21239162.

(3) Frolov, A.; Mamontova, T.; Ihling, C.; Lukasheva, E.; Bankin, M.; Chantseva, V.; Vikhnina, M.; Soboleva, A.; Shumilina, J.; Osmolovskaya, N.; Zhukov, V.; Hoehenwarter, W.; Sinz, A.; Tikhonovich, I.; Wessjohann, L. A.; Bilova, T.; Smolikova, G.; Medvedev, S. Mining Seed Proteome: From Protein Dynamics to Modification Profiles. *Biol. Commun.* **2018**, *63* (1), 43–58.

(4) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359–362. https://doi.org/10.1038/nmeth.1322.

(5) Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D. R.; Omasits, U.; Wollscheid, B.; Volkman, B. F.; Noon, K. R.; Gundry, R. L. Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins. *Anal. Chem.* **2014**, *86* (3), 1551–1559. https://doi.org/10.1021/ac403185a.

(6) Wiśniewski, J. R. Filter Aided Sample Preparation – A Tutorial. *Anal. Chim. Acta* **2019**, *1090*, 23–30. https://doi.org/10.1016/j.aca.2019.08.032.

(7) Erde, J.; Loo, R. R. O.; Loo, J. A. Enhanced FASP (EFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments. *J. Proteome Res.* **2014**, *13* (4), 1885–1895. https://doi.org/10.1021/pr4010019.

(8) Wiśniewski, J. R.; Mann, M. Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis. *Anal. Chem.* **2012**, *84* (6), 2631–2637. https://doi.org/10.1021/ac300006b.

(9) Chong, J.; Wishart, D. S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinforma.* **2019**, *68* (1).

https://doi.org/10.1002/cpbi.86.

(10) Wang, W. Q.; Jensen, O. N.; Møller, I. M.; Hebelstrup, K. H.; Rogowska-Wrzesinska, A. Evaluation of Sample Preparation Methods for Mass Spectrometry-Based Proteomic Analysis of Barley Leaves. *Plant Methods* **2018**, *14* (1), 72. https://doi.org/10.1186/s13007-018-0341-4.

(11) Min, C. W.; Gupta, R.; Agrawal, G. K.; Rakwal, R.; Kim, S. T. Concepts and Strategies of Soybean Seed Proteomics Using the Shotgun Proteomics Approach. *Expert Review of Proteomics*. Taylor and Francis Ltd 2019, pp 795–804. https://doi.org/10.1080/14789450.2019.1654860.

(12) Jiang, Q.; Li, X.; Niu, F.; Sun, X.; Hu, Z.; Zhang, H. ITRAQ-Based Quantitative Proteomic Analysis of Wheat Roots in Response to Salt Stress. *Proteomics* **2017**, *17* (8). https://doi.org/10.1002/pmic.201600265.

(13) Szymanski, J.; Levin, Y.; Savidor, A.; Breitel, D.; Chappell-Maor, L.; Heinig, U.; Töpfer, N.; Aharoni, A. Label-Free Deep Shotgun Proteomics Reveals Protein Dynamics during Tomato Fruit Tissues Development. *Plant J.* **2017**, *90* (2), 396–417. https://doi.org/10.1111/tpj.13490.

(14) Heyer, R.; Schallert, K.; Büdel, A.; Zoun, R.; Dorl, S.; Behne, A.; Kohrs, F.; Püttker, S.; Siewert, C.; Muth, T.; Saake, G.; Reichl, U.; Benndorf, D. A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics within 24 h Using Phenol Extraction, Fasp Digest, and the Metaproteomeanalyzer. *Front. Microbiol.* **2019**, *10* (AUG). https://doi.org/10.3389/fmicb.2019.01883.

(15) Smolko, L.; Smolková, R.; Samoľová, E.; Morgan, I.; Saoud, M.; Kaluđerović, G. N. Two Isostructural Co(II) Flufenamato and Niflumato Complexes with Bathocuproine:

Analogues with a Different Cytotoxic Activity. *J. Inorg. Biochem.* **2020**, *210*. https://doi.org/10.1016/j.jinorgbio.2020.111160.

(16) Bilova, T.; Lukasheva, E.; Brauch, D.; Greifenhagen, U.; Paudel, G.; Tarakhovskaya, E.; Frolova, N.; Mittasch, J.; Balcke, G. U.; Tissier, A.; Osmolovskaya, N.; Vogt, T.; Wessjohann, L. A.; Birkemeyer, C.; Milkowski, C.; Frolov, A. A Snapshot of the Plant Glycated Proteome: Structural, Functional, and Mechanistic Aspects. *J. Biol. Chem.* **2016**, *291* (14), 7621–7636. https://doi.org/10.1074/jbc.M115.678581.

(17) Antonova, K.; Vikhnina, M.; Soboleva, A.; Mehmood, T.; Heymich, M.-L.; Leonova, T.; Bankin, M.; Lukasheva, E.; Gensberger-Reigl, S.; Medvedev, S.; Smolikova, G.; Pischetsrieder, M.; Frolov, A. Analysis of Chemically Labile Glycation Adducts in Seed Proteins: Case Study of Methylglyoxal-Derived Hydroimidazolone 1 (MG-H1). *Int. J. Mol. Sci.* **2019**, *20* (15), 3659. https://doi.org/10.3390/ijms20153659.

(18) Bassal, M.; Abukhalaf, M.; Majovsky, P.; Thieme, D.; Herr, T.; Ayash, M.; Tabassum, N.; Al Shweiki, M. R.; Proksch, C.; Hmedat, A.; Ziegler, J.; Lee, J.; Neumann, S.; Hoehenwarter, W. Reshaping of the Arabidopsis Thaliana Proteome Landscape and Co-Regulation of Proteins in Development and Immunity. *Mol. Plant* **2020**, *13* (12), 1709–1732. https://doi.org/10.1016/j.molp.2020.09.024.

(19) Mamontova, T.; Afonin, A. M.; Ihling, C.; Soboleva, A.; Lukasheva, E.; Sulima, A. S.; Shtark, O. Y.; Akhtemova, G. A.; Povydysh, M. N.; Sinz, A.; Frolov, A.; Zhukov, V. A.; Tikhonovich, I. A. Profiling of Seed Proteome in Pea (Pisum Sativum l.) Lines Characterized with High and Low Responsivity to Combined Inoculation with Nodule Bacteria and Arbuscular Mycorrhizal Fungi. *Molecules* **2019**, *24* (8). https://doi.org/10.3390/molecules24081603.

(20) Greifenhagen, U.; Frolov, A.; Blüher, M.; Hoffmann, R. Plasma Proteins Modified by Advanced Glycation End Products (AGEs) Reveal Site-Specific Susceptibilities to Glycemic Control in Patients with Type 2 Diabetes. *J. Biol. Chem.* **2016**, *291* (18), 9610–9616. https://doi.org/10.1074/jbc.M115.702860.

(21) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450. https://doi.org/10.1093/nar/gky1106.

(22) Potriquet, J.; Laohaviroj, M.; Bethony, J. M.; Mulvenna, J. A Modified FASP Protocol for High-Throughput Preparation of Protein Samples for Mass Spectrometry. *PLoS One* **2017**, *12* (7), e0175967. https://doi.org/10.1371/journal.pone.0175967.

(23) Ni, M.; Wang, L.; Chen, W.; Mou, H.; Zhou, J.; Zheng, Z. Modified Filter-Aided Sample Preparation (FASP) Method Increases Peptide and Protein Identifications for Shotgun Proteomics. *Rapid Commun. Mass Spectrom.* **2017**, *31* (2), 171–178. https://doi.org/10.1002/rcm.7779.

(24) Mitchell, P. Proteomics Retrenches. *Nature Biotechnology*. Nature Publishing Group 2010, pp 665–670. https://doi.org/10.1038/nbt0710-665.

(25) Tuli, L.; Tsai, T. H.; Varghese, R. S.; Xiao, J. F.; Cheema, A.; Ressom, H. W. Using a Spike-in Experiment to Evaluate Analysis of LC-MS Data. *Proteome Sci.* **2012**, *10* (1), 13. https://doi.org/10.1186/1477-5956-10-13.

(26) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative Proteomic Analysis

of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **2012**, *11* (3), 1–11. https://doi.org/10.1074/mcp.M111.014050.

(27) Beri, J.; Rosenblatt, M. M.; Strauss, E.; Urh, M.; Bereman, M. S. Reagent for Evaluating Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) Performance in Bottom-Up Proteomic Experiments. *Anal. Chem.* **2015**, *87* (23), 11635–11640. https://doi.org/10.1021/acs.analchem.5b04121.

(28) Chang, Y. H.; Lee, S. H.; Liao, I. C.; Huang, S. H.; Cheng, H. C.; Liao, P. C. Secretomic Analysis Identifies Alpha-1 Antitrypsin (A1AT) as a Required Protein in Cancer Cell Migration, Invasion, and Pericellular Fibronectin Assembly for Facilitating Lung Colonization of Lung Adenocarcinoma Cells. *Mol. Cell. Proteomics* **2012**, *11* (11), 1320–1339. https://doi.org/10.1074/mcp.M112.017384.

(29) Zauber, H.; Schüler, V.; Schulze, W. Systematic Evaluation of Reference Protein Normalization in Proteomic Experiments. *Front. Plant Sci.* **2013**, *4* (FEB), 25. https://doi.org/10.3389/fpls.2013.00025.

(30) Isaacson, T.; Damasceno, C. M. B.; Saravanan, R. S.; He, Y.; Catala, C.; Rose, J. K. C. Sample Extraction Techniques for Enhanced Proteomic Analysis of Plant Tissues. *Nat. Protoc.* **2006**, *1* (2), 769–774. https://doi.org/10.1038/nprot.2006.102.

(31) Mamontova, T.; Lukasheva, E.; Mavropolo-Stolyarenko, G.; Proksch, C.; Bilova, T.; Kim, A.; Babakov, V.; Grishina, T.; Hoehenwarter, W.; Medvedev, S.; Smolikova, G.; Frolov, A. Proteome Map of Pea (Pisum Sativum L .) Embryos Containing Different Amounts of Residual Chlorophylls. *Int. J. Mol. Sci.* **2018**, *19*, 1–20. https://doi.org/10.3390/ijms19124066.

Graphical TOC