JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

Cite This: J. Agric. Food Chem. 2019, 67, 10296−10305

pubs.acs.org/JAFC

Proteomic Comparison of Three Extraction Methods Reveals the Abundance of Protease Inhibitors in the Seeds of Grass Pea, a Unique Orphan Legume

Carrie Miranda,[†,⊥] Quanle Xu,[‡,⊥] Nathan W. Oehrle,[†] Nazrul Islam,[§] Wesley M. Garrett,[∥] Savithiry S. Natarajan,[§] Jason D. Gillman,[†] and Hari B. Krishnan*[,†]

[†]Plant Genetics Research, USDA-Agricultural Research Service, University of Missouri, Columbia, Missouri 65211, United States

[‡]College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China

[§]Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, Maryland 20705, United States

[∥]Animal Bioscience and Biotechnology Laboratory, USDA-Agricultural Research Service, Beltsville 20705, United States

Supporting Information

ABSTRACT: Grass pea is an orphan legume that is grown in many places in the world. It is a high-protein, drought-tolerant legume that is capable of surviving extreme environmental challenges and can be a sole food source during famine. However, grass pea produces the neurotoxin *β*-*N*-oxalyl-L-*α*,*β*-diaminopropionic acid (*β*-ODAP), which can cause a neurological disease. This crop is promising as a food source for both animals and humans if *β*-ODAP levels and other antinutritional factors such as protease inhibitors are lowered or removed. To understand more about these proteins, a proteomic analysis of grass pea was conducted using three different extraction methods to determine which was more efficient at isolating antinutritional factors. Seed proteins extracted with Tris-buffered saline (TBS), 30% ethanol, and 50% isopropanol were identified by mass spectrometry, resulting in the documentation of the most abundant proteins for each extraction method. Mass spectrometry spectral data and BLAST2GO analysis led to the identification of 1376 proteins from all extraction methods. The molecular function of the extracted proteins revealed distinctly different protein functional profiles. The majority of the TBS-extracted proteins were annotated with nutrient reservoir activity, while the isopropanol extraction yielded the highest percentage of endopeptidase proteinase inhibitors. Our results demonstrate that the 50% isopropanol extraction method was the most efficient at isolating antinutritional factors including protease inhibitors.

KEYWORDS: *grass pea, proteomics, storage proteins, protease inhibitors, β-ODAP*

INTRODUCTION

Grass pea (*Lathyrus sativus L.*) is a drought-tolerant, high-protein legume that is grown for human consumption or forage in Europe, Asia, and East Africa.[1] It is tolerant of many harsh abiotic stresses, including drought and flooding, because of its resilient root system and has become a staple crop in times of famine and difficult climatic conditions.[2−6] In addition, the seed is an excellent source of plant protein, containing 25−29% protein on a dry weight basis, and as a legume, it is capable of nitrogen fixation.[7,8] Grass pea is a part of many human-consumed food dishes in Spain, India, Bangladesh, and especially Ethiopia.[9,10]

Although grass pea has nutritional potential for semiarid and arid places in the world, its production is hindered because of an antinutritional factor *β*-*N*-oxalyl-L-*α*,*β*-diaminopropionic acid (*β*-ODAP) in its seeds.[11−13] When the seeds are grown under water-stressed conditions, accumulation of *β*-ODAP increases.[13,14] When grass pea is consumed in large quantities for a long period of time without other food sources in the diet, such as during a famine, a disease called lathyrism develops, which affects both humans and animals.[10,12,15] Lathyrism is a neurological disease which can cause permanent paralysis and brain damage, especially in children.[16] A famine in Ethiopia in the 1970s left 1% of the population disabled because of the dependence on grass pea.[15] In addition to *β*-ODAP, other antinutritional factors exist in grass pea, including protease inhibitors, oligosaccharides, tannins, and phytic acids.[17−20] The two most abundant protease inhibitors in grass pea seeds are trypsin inhibitors and Bowman−Birk inhibitors.[21] Trypsin inhibitors inhibit the mammalian digestive enzyme trypsin, and Bowman−Birk inhibitors similarly affect the serine peptidase digestive enzymes.[22] Protease inhibitors, when not thoroughly denatured by cooking, can interfere with the tryptic breakdown of protein and are abundant in legumes.[22] There have been breeding efforts in several legume species, including soybean, to reduce trypsin inhibitor levels for human health and animal feed conversion efficiency. Breeding efforts have been initiated aiming at developing new *L. sativus* varieties that have reduced antinutritional factors.[8] Although detailed knowledge of the grass pea seed proteins would be invaluable to improve breeding efforts, a proteomic analysis of grass pea seeds has not previously been conducted. Storage proteins make up the most abundant proteins in a seed, which can overshadow lower

Received: July 9, 2019
Revised: August 27, 2019
Accepted: August 29, 2019
Published: August 29, 2019

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

expressed but more impactful proteins, such as several proteinaceous, antinutritional factors.[23] Various modified protein extraction methods have been developed, which can limit the solubility of the main storage proteins (e.g., globulins) allowing lower abundance proteins to be enriched, detected, and quantified.[23−26] Here, we introduce the first proteomic study of grass pea seeds. We conducted and evaluated the performance of three separate protein extractions in order to (1) quantify the protein profiles of the most abundant proteins and (2) test two different alcohol precipitation methods to enrich less abundant seed proteins. We then characterized all protein profiles based on their predicted gene ontologies (cellular location, molecular function, and biological process). Finally, we analyzed the most abundant proteins for each extraction and made a direct comparison of the specific proteins enriched using each method. These results suggest that a 50% isopropanol extraction is more efficient at isolating protease inhibitors than the commonly applied Tris-buffered saline (TBS) and an alternate 30% ethanol extraction method.

MATERIALS AND METHODS

Seed Protein Extraction. Dry seeds of *L. sativus* cv. LZ(2), a Chinese cultivar, were exclusively used in this study. The seeds were ground to a fine powder with a mortar and pestle. For one-dimensional (1D) sodium dodecyl sulfate-polyacrylamide gel analysis (SDS-PAGE), aliquots of seed powder (50−100 mg) were transferred to 2 mL microcentrifuge tubes and separately extracted either with 1 mL of TBS (50 mM Tris-Cl, pH 7.5, 150 mM NaCl) or 30% ethanol (v/v) or 50% isopropanol (v/v). The content of the tubes were vigorously vortexed for 15 min at room temperature, followed by centrifugation at 15 800×*g* for 5 min. The clear supernatant was collected and three volumes of ice-cold acetone were added. The contents were mixed and left at −20 °C overnight. Precipitated proteins were recovered by centrifugation at 15 800×*g* for 5 min. The protein pellet was briefly air-dried and resuspended in SDS sample buffer [60 mM Tris-HCl, pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), and 5% 2-mercaptoethanol (v/v)] for 1D gel analysis.

For two-dimensional (2D) gel analysis, acetone-precipitated proteins from the three extractions (TBS, 30% ethanol, and 50% isopropanol) were individually suspended in 5 mL of 100 mM Tris-HCl, pH 8.8, 0.9 M sucrose, and 0.4% 2-mercaptoethanol (v/v). To extract only the proteins from these suspensions and remove the interfering compounds, an equal volume of Tris-saturated phenol (pH 4.3) was added, and the contents were mixed vigorously for 30 min at 25 °C. Subsequently, they were subjected to centrifugation at 5000×*g* for 20 min at 25 °C in a swing-bucket rotor. The upper phenolic phase was removed and added to 10 volumes of freshly prepared 100% methanol with 0.1 M ammonium acetate (chilled to −80 °C). Precipitation of the extracted proteins progressed for 2 h at −80 °C, followed by centrifugation at 12 000×*g* for 15 min at 4 °C. The supernatant was discarded, and the protein pellet was suspended vigorously in a freshly prepared solution of 100% methanol with 0.1 M ammonium acetate and 0.01 M dithiothreitol (DTT) chilled to −20 °C. Washing of the insoluble proteins was repeated three times with the same solution with incubation at −20 °C for 20 min, followed by centrifugation at 12 000×*g* for 10 min at 4 °C between each vigorous wash step. Washing of the insoluble proteins was repeated four more times with a freshly prepared solution of 100% acetone containing 0.01 M DTT with incubation at −20 °C for 20 min, followed by centrifugation at 12 000×*g* for 10 min at 4 °C between each vigorous wash step. After the final centrifugation, the protein pellet was allowed to air-dry slightly and then solubilized in a small volume of 7 M urea, 2 M thiourea, 1% (w/v) of 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, and 2% (w/v) 3-(4-heptyl)phenyl-3-hydroxypropyl)dimethyl-ammoniopropanesulfonate (C7BzO). The samples were placed on ice, and the protein concentration was estimated following the method of Bradford.

1D and 2D PAGE Analysis of Grass Pea Proteins. The 1D and 2D gel electrophoreses were performed as described earlier.[26]

Mass Spectrometry. Grass pea seed proteins for mass spectrometry were extracted following the same procedure as for 2D gels. Insoluble proteins, once in acetone (this time without DTT), were washed three times in 100% acetone. The proteins were then suspended in 500 μL of lysis buffer containing 6 M urea and 100 mM TBS-HCl (pH 7.8). The protein concentration was determined by the bicinchoninic acid method (Pierce, Rockford, IL, USA). Proteins (∼200 μg) were reduced in TBS (2-carboxyethyl) phosphine, carbamidomethylated with iodoacetamide, and digested overnight at 37 °C with trypsin at the ratio of 100:1. The digested peptides were purified by reverse-phase chromatography using SEP-PAK C18 columns (Waters Corp, Milford, MA, USA). The peptides were suspended in 50 μL of 5% acetonitrile (v/v) and 0.2% formic acid (v/v). The peptide concentration was determined by the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Waltham, MA, USA).

Protein identification by mass spectrometry was performed as described previously.[27] In brief, peptides (∼500 ng per fraction) were separated on a 75 μm (inner diameter) fused silica capillary pulled to a 5 micron tip with a P-2000 Sutter laser puller (Sutter Instrument, Novato, CA, USA), packed with 22 cm of 2.5 μm Synergi Hydro-RP C18 (Phenomenex, Torrance, CA, USA), and coupled directly to a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) controlling a 120 min linear gradient from 3.2 to 40% acetonitrile and 0.1% formic acid at a flow rate of 300 nL per minute. Peptides were electrosprayed at 2.4 kV into an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) operating in data-dependent mode with positive polarity and using 445.12003 m/z as an internal mass calibrant. Quadrupole isolation was enabled, and survey scans were recorded in an Orbitrap at 120 000 resolution over a mass range of 400−1600 m/z. The instrument was operated in top speed mode with a cycle time of 3 s. Peptides were isolated using the quadrupole with an isolation window of 1.6 m/z and then fragmented by high-energy collision-induced dissociation in the ion-routing multipole and the resulting fragment ions were detected in the linear ion trap.

Peptide Matching. Raw data files were converted into searchable peak lists and saved as mascot generic format (.mgf) files with RawConverter.[28] Mass spectrometry data were searched with Mascot 2.5.1 (Matrix Science, London, UK). Because *L. sativus* has minimal sequence coverage in public protein databases, a homology-based search strategy was employed for protein identification. To accomplish this, a multispecies legume-specific FASTA-formatted database was constructed by downloading the protein reference sequence subsets from the National Center for Biotechnology Information nonredundant protein database using the taxonomy filter Fabales. Common contaminants were appended to this database for a total of 483 721 sequences as of 02/08/2018. These data were simultaneously searched against the Swissprot database (12/2017 version) using the taxonomy filter, other green plants, which resulted in 18 876 additional sequences. The Mascot search parameters were as follows: monoisotopic mass; parent ion tolerance of 5 ppm; fragment ion tolerance of 0.6 Da; ^{13}C isotopes set to 2; peptide charge states of 1+, 2+, and 3+; trypsin as a digesting enzyme with one missed cleavage allowed; fixed modification of carbamidomethyl C; and variable modifications of oxidation of methionine and N-terminal pyroglutamic acid from glutamic acid or glutamine. A scaffold (version Scaffold_4.8.4, Proteome Software Inc., Portland, OR, USA) was used to validate peptide and protein identifications. Peptide identifications were accepted if they could be established at >95.0% probability as specified by the Peptide Prophet algorithm[29] and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm.[30]

Gene Ontology Analysis. Database matches for each of the three experimental protein extraction methods identified by peptide searches of the SwissProt_2017_12.fasta or refseq_fabales_020817.-fasta databases were downloaded (ethanol, 938 proteins; isopropanol,

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

311 proteins; TBS, 908 proteins; and composite, 1376 proteins) (Table S1: Peptide Database Search). The BLAST2GO[31] plugin of CLC genomic workbench was then used to identify Gene Ontology (GO) slim plant terms associated with each of the proteins identified either by BLAST2GO databases or through the Interpro database using an *E*-value threshold cutoff for BLASTP searches of 1×10^{-5} (Table S1: BLAST2GO GO).

Statistical Analysis. Statistical analysis to generate the heatmap was done with JMP Pro (Version 13.1), using exclusive spectral counts from in-solution digests of the three protein isolation methods.

RESULTS AND DISCUSSION

1D SDS-PAGE Analysis of Grass Pea Seed Proteins Using Three Different Extraction Methods. By mass, the majority of proteins in a legume seed are classified as storage proteins,[32] which typically have no enzymatic function but serve as a repository of nitrogen, carbon, and to a lesser extent energy. The abundance of these proteins is critical for seed viability but can overshadow less abundant, but economically more important, proteins such as seed antinutritional factors. As proteins can be separated and distinguished based on their different solubility properties and molecular weights, these features can be exploited during the extraction process to create different protein profiles. Several different methods of protein extraction have been tested in other crop species to determine the best method to visualize lower abundance proteins for proteomic analysis.[33,34] The protein profiles of soybean and other legume seeds have been characterized well; however, this has not been done for grass pea seeds.[26,35,36] Being able to understand the protein profile of grass pea will assist researchers to identify antinutritional factors, such as protease inhibitors, and proteins that are involved in the synthesis of β-ODAP. First, we explored the protein profile that can be visualized by a 1D SDS-PAGE gel after protein extraction (Figure 1). The most common method of protein extraction is TBS extraction, which produced a large range of proteins between 10 and 97 kDa. Previous studies have demonstrated that this method favors enrichment of storage proteins.[32] In order to detect and quantify less abundant seed proteins, we compared TBS with two alternate protein extraction methods that use either 30% ethanol or 50% isopropanol. We hypothesized that by performing seed protein extractions using buffers with different effects on storage protein solubility, a unique protein profile enriched in nonstorage proteins could be achieved, as was successfully shown in soybean.[37] SDS-PAGE comparison of three protein extraction methods demonstrates that all three extraction methods produced unique protein profiles (Figure 1). The 30% ethanol and 50% isopropanol extractions had a protein profile between 10 and 97 kDa similar to the TBS extraction; however, the intensity and molecular weights of proteins varied within each extraction method. The ethanol extraction showed an abundance of protein(s) with molecular weights at ~50, 40, and 28 kDa, which were not as prominent as in the other extractions. The 50% isopropanol extraction showed large molecular weight proteins only faintly and favored lower molecular weight proteins as compared with the other two extraction methods; however, it appears that fewer proteins were extracted with this method. Thus, the size of the protein, amino acid composition, as well as the nature of the solvent may play an important role in determining the differential solubility of grass pea proteins.

2D SDS-PAGE Analysis of Grass Pea Seed Proteins Using Three Different Extraction Methods. To compare the extracted proteins from each method, 2D SDS-PAGE gel electrophoresis was conducted where proteins were separated by isoelectric point and molecular weight. The pI separation range was pI 4−7, and proteins extracted from the three different methods were visualized by Coomassie gel staining (Figure 2). 2D electrophoresis allowed the separation of grass pea proteins into several discrete protein spots. The TBS extraction shows a protein profile to enrich storage proteins, which was suggested by the appearance of the 1D gel (Figure 2A). The 30% ethanol extraction (Figure 2B) and the 50% isopropanol extraction (Figure 2C) show very different protein profiles. Few proteins are similar between the TBS and ethanol extractions. These gel comparisons suggest that the TBS extraction will favor abundant proteins. The ethanol extraction is capable of eliminating many, but not all, of the most abundant proteins. Finally, the isopropanol extraction produces the most unique protein profile of all tested extraction methods; however, it is unclear which minor proteins are present.

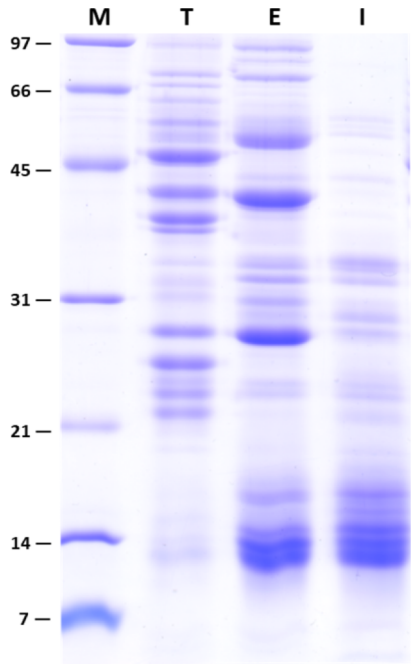

Figure 1. 1D SDS-PAGE of grass pea total seed protein as extracted by TBS (lane T), 30% ethanol (lane E), and 50% isopropanol (lane I). Seed proteins were separated using a SDS−polyacrylamide gel and visualized by gel staining with Coomassie Blue. Molecular weight markers are shown in lane M with kDa sizes on the left.

Identification of Grass Pea Proteins. The identity of all grass pea proteins isolated by TBS, 30% ethanol, and 50% isopropanol extraction methods is shown in Table S1: Summarized Spectrum Data (all), TBS Summarized Spectrum Data, ISO Summarized Spectrum Data, and EtOH Summarized Spectrum Data. Proteins were identified as a composite of all three extraction methods and also through individual methods. Proteins were identified via searches of mass spectral data against peptide databases NCBI and Swissprot, which utilized amino acid sequence data from multiple species to predict protein identity. The most abundant in the TBS extraction were storage proteins, accounting for 64% of all extracted proteins. The TBS method was able to isolate both water-soluble and salt-soluble proteins (Table S1: TBS Summarized Spectrum Data). Storage proteins were also

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

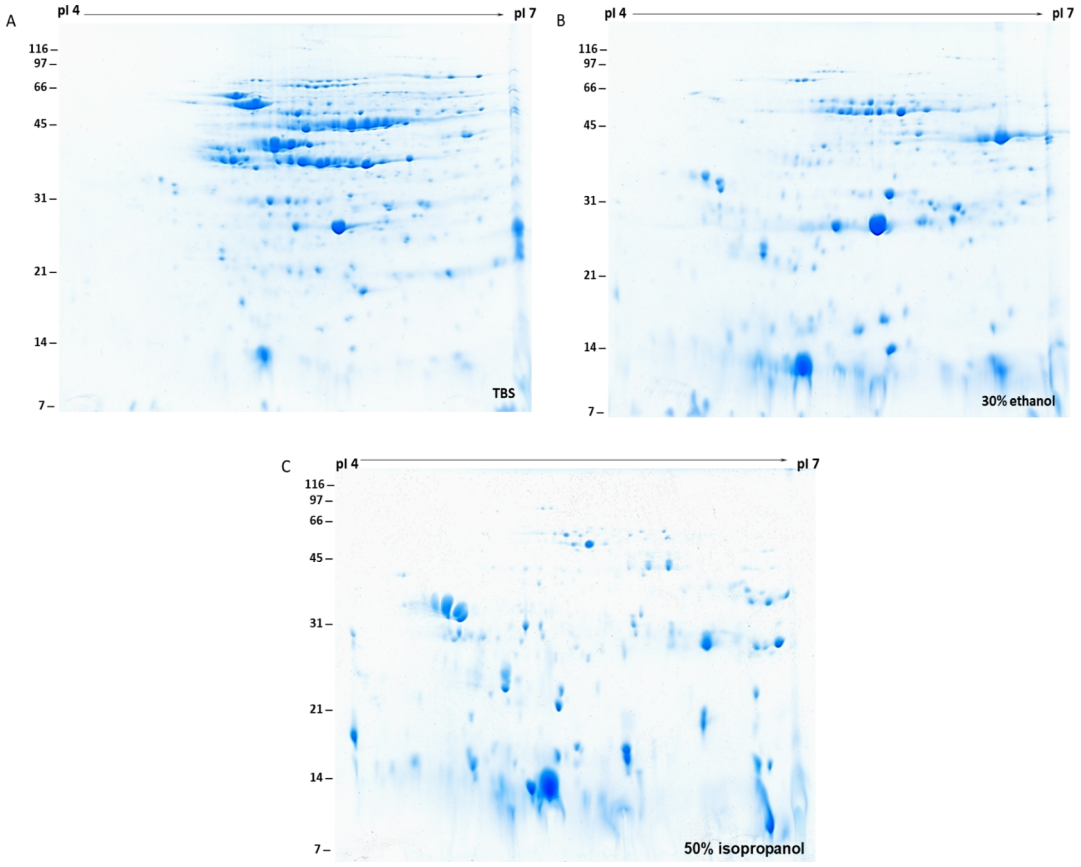

Figure 2. 2D gel electrophoresis comparison of three protein extraction methods. Proteins (300 μg) from three protein extraction methods were separated by isoelectric focusing on pI 4−7 strips, followed by SDS-PAGE on 15% gels. Following electrophoresis, the gels were stained with Colloidal Coomassie Blue G-250. The position and sizes of protein markers in kDa are shown on the left side of the figures. A. TBS extraction; B. 30% ethanol extraction; and C. 50% isopropanol extraction.

abundant in the ethanol and isopropanol extraction methods accounting for 28 and 20%, respectively, of all proteins.

Each extraction method was able to isolate different subsets of total proteins in dissimilar quantities as the banding/spot patterns observed in Figures 1 and 2 suggest. In the TBS extraction method, a total of 908 proteins were identified; the most abundant at 10% was the storage protein legumin (Table S1: TBS Summarized Spectrum Data). In the 30% ethanol extraction, 938 proteins were identified, and the most abundant protein (5% of total spectral counts) was the Bowman−Birk protease inhibitor, followed by storage albumin proteins (Table S1: EtOH Summarized Spectrum Data). In the 50% isopropanol extraction, only 311 proteins were identified, but the majority (21.6% of total spectral counts) was the Bowman−Birk protease inhibitor (Table S1: ISO Summarized Spectrum Data).

Table 1 lists the most abundant proteins from each extraction (except storage proteins). This analysis will allow for easier comparison of less abundantly expressed proteins from all three extraction methods. Full details on proteins identified are found in Table S1: Summarized Spectrum Data (all). Aside from the storage proteins, the most abundant protein isolated from all three extraction methods are Bowman−Birk-type protease inhibitors (Table 1). The TBS extraction method was able to isolate proteins that bind biotin and are believed to play a role in cell growth.[38] Subsequent to that are heat shock proteins, which, similar to late embryogenesis proteins, help to protect the embryo from desiccation.[39,40] Fructose-bisphosphate aldolase is also isolated and is necessary for glycolysis. 1-Cys peroxiredoxin is a regulator of seed dormancy, inhibiting germination during stress.[41]

The ethanol extraction method was efficient in isolating alcohol dehydrogenase (an enzyme that protects embryos from oxidative stress), fructose-biphosphate aldolase, and heat shock proteins as well. Interestingly, it was the only method that was able to detect cysteine synthase in the top 30 proteins (Table 1), an enzyme necessary for catalyzing the reaction: O_3-acetyl-L-serine + hydrogen sulfide ↔ L-cysteine + acetate. Cysteine is a semiessential sulfur-containing amino acid, which must be ingested when a diet is insufficient in methionine.[42] Cysteine also plays a role in the β-ODAP synthesis pathway as a precursor to β-ODAP.[4] The ethanol extraction was the only method to detect cyanoalanine synthase (Table S1: EtOH Summarized Spectrum Data, rows 890 and 936). This enzyme also plays a key role in the β-ODAP synthesis by generating precursors to β-ODAP through two different reactions, one using cysteine and another using isoxazolin-5-one.[4]

In the isopropanol extraction method, the Bowman−Birk protease inhibitor, an antinutritional factor, was the most abundant proteins isolated (Table 1). Following that were superoxide dismutases and glutaredoxins, which protect seed embryos from oxidative stress.[43,44] An abundant protein in this extraction is the SLE1 protein that serves a similar function as late embryogenesis proteins[45] and is also common in the other two extractions.

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

Table 1. List of Names and Exclusive Spectral Count (ct) of Identified Proteins from Three Different Methods

TBS biological sample name	Ct	30% EtOH biological sample name	Ct	50% ISO Biological sample name	Ct
Bowman−Birk-type proteinase inhibitor 1	349	Bowman−Birk-type proteinase inhibitor 1	270	Bowman−Birk-type proteinase inhibitor 1	515
seed biotin-containing protein	64	Fructose-bisphosphate aldolase, cytoplasmic isozyme 2	108	superoxide dismutase	69
heat shock 70 kDa protein	46	heat shock 70 kDa protein	95	glutaredoxin C4	69
fructose-bisphosphate aldolase	33	alcohol dehydrogenase 1	85	calmodulin-1/11/16	65
PREDICTED: 1-Cys peroxiredoxin isoform X1	31	phosphopyruvate hydratase	84	protein SLE1	41
glyceraldehyde-3-phosphate dehydrogenase	30	glyceraldehyde-3-phosphate dehydrogenase	67	PREDICTED: late embryogenesis abundant protein	36
phosphopyruvate hydratase	26	protein SLE1	53	phosphopyruvate hydratase	30
alpha-1,4 glucan phosphorylase L isozyme	25	ABA-responsive protein ABR18	46	seed biotin-containing protein SBP65	28
alcohol dehydrogenase 1	24	glutaredoxin C4	45	plant/MUD21-2 protein	28
protein SLE1	21	late embryogenesis abundant protein B19.1A	38	PREDICTED: glutaredoxin	21
seed linoleate 9S-lipoxygenase-3	18	superoxide dismutase [Cu−Zn]	34	MFP1 attachment factor-like protein	21
glutaredoxin C4	18	aldo/keto reductase family oxidoreductase	33	ABA-responsive protein ABR18	20
PREDICTED: late embryogenesis abundant protein D-34-like isoform X2	18	nucleoside diphosphate kinase 1	31	PREDICTED: 1-Cys peroxiredoxin isoform X1	19
PREDICTED: elongation factor 1-alpha-like	18	alpha-1,4 glucan phosphorylase L isozyme, chloroplastic/amyloplastic	31	import inner membrane translocase protein	18
dehydrin-like protein	17	malate dehydrogenase	30	nucleoside diphosphate kinase 1	17
malate dehydrogenase	17	PREDICTED: 1-Cys peroxiredoxin isoform X1	28	peptidyl-prolyl cis−trans isomerase FKBP12	16
translation elongation factor EF-2 subunit	16	20S proteasome subunit alpha	27	PREDICTED: 10 kDa chaperonin	15
nucleoside diphosphate kinase 1	15	seed linoleate 9S-lipoxygenase-3	26	lectin alpha-1 chain	14
superoxide dismutase [Cu−Zn]	15	aldehyde dehydrogenase family 7 member A1	26	NAD(P)-binding Rossmann-fold protein	14
60S ribosomal L12-like protein	15	hypothetical protein PHAVU_001G067300g	26	PREDICTED: huntingtin-interacting protein K-like	14
ABA-responsive protein ABR18	14	20S proteasome alpha subunit E1	23	EF hand calcium-binding family protein	13
PREDICTED: phosphoglycerate kinase, cytosolic	14	PREDICTED: late embryogenesis abundant protein D-34-like	20	subtilisin inhibitor 1	12
F0F1-type ATP synthase, beta subunit	14	PREDICTED: proteasome subunit alpha type-6	20	2-dehydro-3-deoxyphosphooctonate aldolase	11
histone H2B	12	PREDICTED: serine hydroxymethyltransferase 4	20	phospholipid hydroperoxide glutathione peroxidase	11
GTP-binding nuclear protein ran/TC4	12	dehydrin-like protein	19	malate dehydrogenase	11
outer plastidial membrane protein porin	12	lactoylglutathione lyase	19	prefoldin	11
60S ribosomal L8-like protein	12	ubiquitin-conjugating enzyme	19	ATP synthase subunit delta', mitochondrial	10
26S protease regulatory subunit 6A homolog	12	Lectin OS = glycine max GN = LE1 PE = 1 SV = 1	18	prefoldin subunit 5	10
14-3-3-like protein A	11	PREDICTED: adenylate kinase 4	18	PREDICTED: protein CutA, chloroplastic isoform X1	10
GTP-binding elongation factor Tu family protein	11	cysteine synthase	12	cytochrome B-c1 complex subunit 6	9

J. Agric. Food Chem. 2019, 67, 10296−10305
DOI: 10.1021/acs.jafc.9b04307

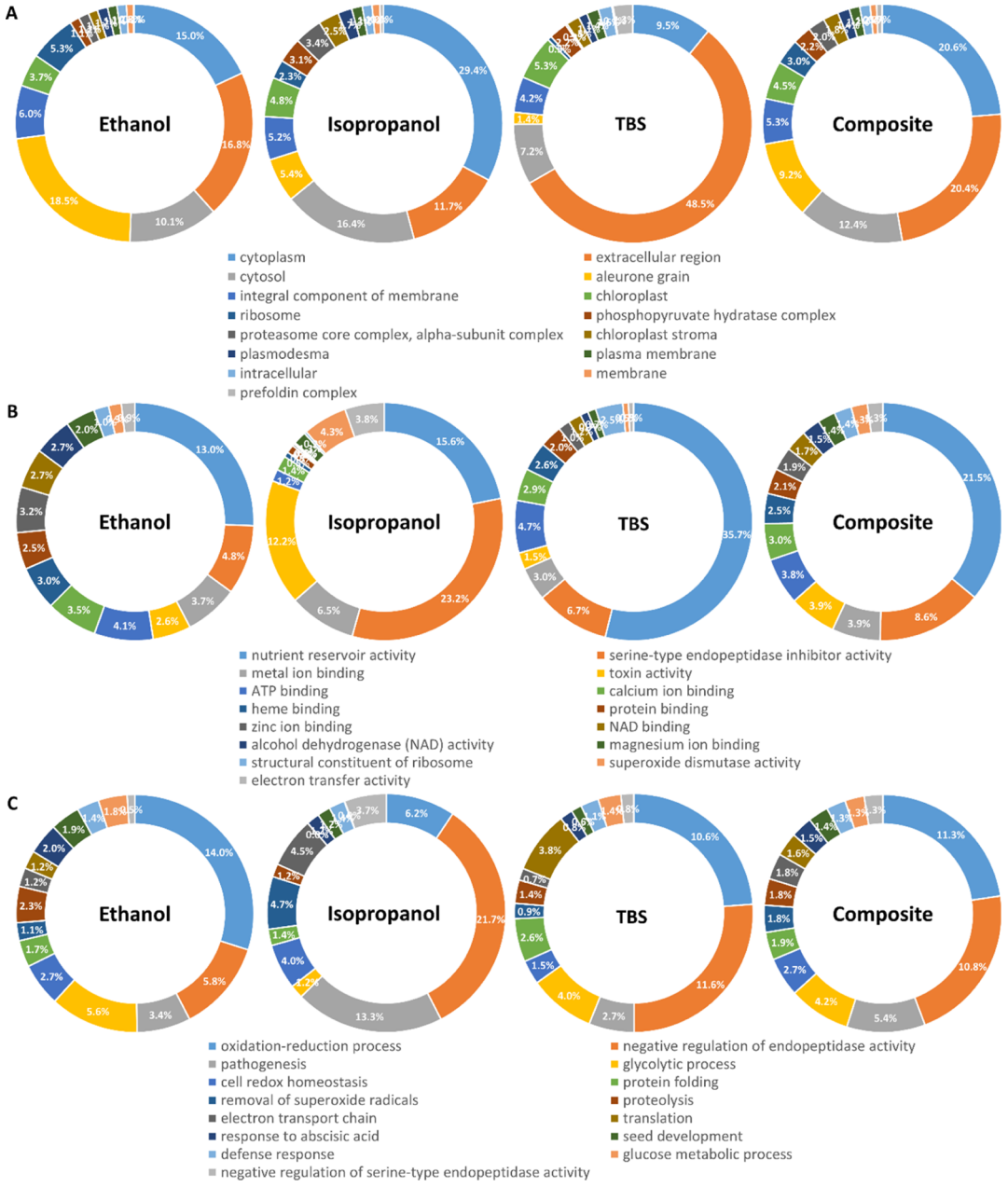

Figure 3. Comparative distribution of top 15 GOSlim Plant terms present in proteins isolated from grass pea seeds using three different protein extraction methods. (A) Cell compartment, (B) molecular function, and (C) biological process.

Functional Classification of Grass Pea Seed Proteins. To annotate protein matches with gene function, the program BLAST2GO[31] was utilized. This software allows for prediction of protein cellular location, molecular function, and biological process by identifying homologous proteins with known functions/localization from other species and inferring the function. It was necessary to rely on protein annotation in species other than grass pea, as protein data for grass pea are extremely limited, and grass pea does not have a complete sequenced genome. Figure S1 shows the species distribution used for protein prediction for grass pea. In the model legume *Medicago*, the protein database is highly annotated and thus searches tend to reveal *Medicago* matches because of the abundance of proteomic data for this species within NCBInr. BLAST2GO results are typically biased toward abundance of species sequences in databases and are not representative of relatedness. For example, both *Pisum* and *Lathyrus* are nested within the tribe Fabeae;[46−48] however, functional proteomic data of *Pisum* and *Lathyrus* are sparse. Had *Pisum* been annotated to the level of *Medicago* or even *Glycine max*, it could be conceivable that most matches would be to that species.

After protein identification, the annotation information for each protein was inferred from several databases (Figure S2). To limit these data to the most important, a comparative distribution of GOSlim Plant terms was performed for cellular location, molecular function, and the biological process of extracted proteins. Figure 3A shows the cellular location of all proteins from each extraction method and a composite analysis, which revealed the three extraction methods each having unique protein profiles. In the TBS extraction, nearly half of the proteins were annotated as corresponding to the

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

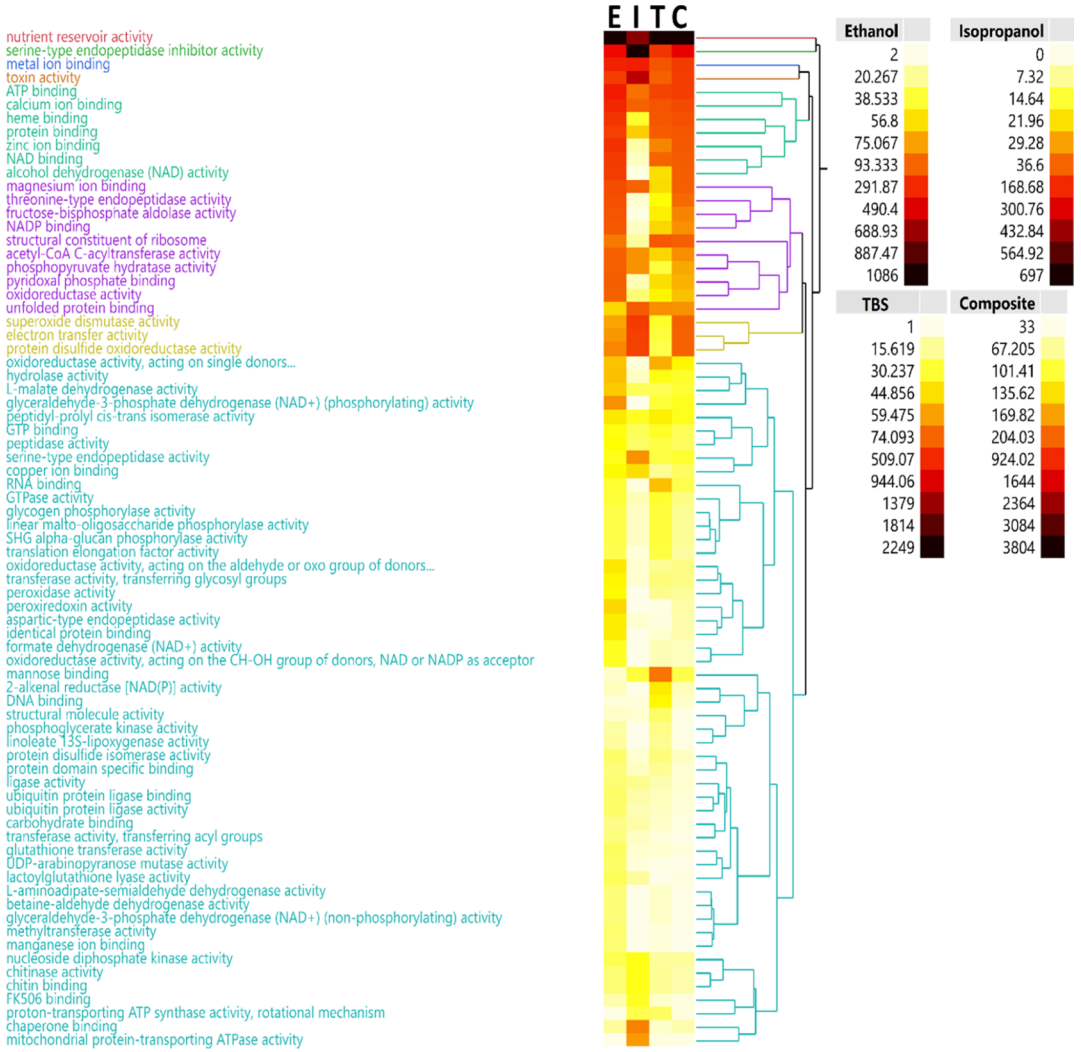

Figure 4. Hierarchical clustering and heatmap of top 75 most abundant molecular function GOSlim terms in proteins isolated using three different isolation methods from grass pea seeds; "E" corresponds to 50% ethanol, "I" corresponds to 30% isopropanol, "T" corresponds to TBS isolation, and "C" corresponds to the composite data from all three methods. Heatmap scale indicates the cumulative exclusive spectral protein count for GOSlim terms.

extracellular region, which is typically rich in storage proteins. The ethanol extraction had nearly equal amounts of proteins annotated as targeted to cytoplasmic, extracellular, and aleurone grain. The isopropanol extraction has a majority of cytoplasmic proteins (Figure 3A). The molecular function of the extracted proteins was also analyzed (Figure 3B) and revealed distinctly different protein functional profiles. The majority of the TBS-extracted proteins are annotated with "nutrient reservoir activity", which is analogous to "storage protein". The ethanol-extracted proteins had more or less equal distribution of protein annotations, with the exception of nutrient reservoir-active proteins which were present ~2.5× higher than the next category which were serine-type endopeptidase inhibitors or simply protease inhibitors. The isopropanol extraction yielded the highest percentage of proteinase inhibitors compared to the other two methods, suggesting that this method is favorable to specifically enrich extraction of protease inhibitors. Interestingly, the isopropanol extraction also has a 6× higher extraction of proteins annotated as "toxins" as compared to the other two methods (Figure 3B). Figure 3C examines the biological process in which the extracted proteins are involved. The majority of proteins extracted by the TBS method either play a role in the oxidation−reduction process or negatively regulate endopeptidases (i.e., protease inhibitors). In the 30% ethanol extraction, the majority of proteins were annotated as playing a role in the oxidation−reduction process. The second most plentiful proteins either negatively regulate endopeptidases (i.e., protease inhibitors) or play a role in the glycolytic process. In the 50% isopropanol extraction, ~22% of the proteins are endopeptidase proteinase inhibitors, followed by proteins that play a role in pathogenesis. These findings indicate that isopropanol extraction is successful at isolating protease inhibitors as well as proteins involved in plant defenses. 13.3% of proteins in the isopropanol extraction were determined to have toxic/pathogenic roles compared to 3.4% in the ethanol extraction and only 2.7% on the TBS extraction.

Heatmap and Clustering of the Most Abundant Molecular Function of Grass Pea Proteins. A more detailed protein expression level analysis for the molecular functions of extracted proteins is shown in Figure 4. Hierarchical clustering was performed for the 75 most abundant molecular functions, and a heatmap display was created showing cumulative exclusive spectral counts (Figure

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

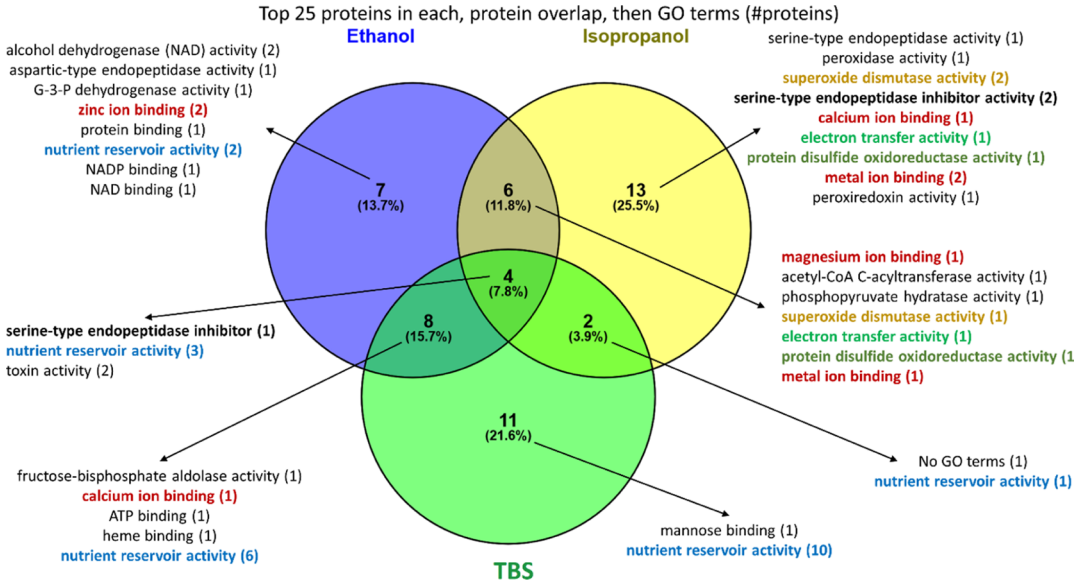

Figure 5. Venn diagram showing top 25 most abundant protein database matches in protein isolates from grass pea seeds using three different protein extraction methods. GOSlim terms present in each category are indicated, along with the number of proteins identified in parentheses. Functional classifications present in multiple overlaps are indicated by colored text. Venn diagrams were created using an online tool http://bioinfogp.cnb.csic.es/tools/venny/index.html.

4). The molecular functions of proteins extracted by TBS, 30% ethanol, 50% isopropanol, and a composite representing all three methods are shown. The TBS and ethanol extraction showed similar results for the most abundant protein function which was nutrient reservoir-active, or storage protein; however, they were not the same specific storage proteins. The TBS extraction favored the legumin protein, but the ethanol method extracted the albumin protein [Table S1: Summarized Spectrum Data (all)]. The most abundant molecular function present in proteins isolated by the isopropanol extraction is the serine-type endopeptidase inhibitor activity, specifically the Bowman−Birk-type protease inhibitors (Table S1: ISO Summarized Spectrum Data, Figures 3 and 4). The second most abundant protein molecular function corresponded to the nutrient reservoir activity, and albumin protein, similar to the ethanol extraction (Table S1: ETOH Summarized Spectrum Data). Proteins annotated as having nutrient reservoir activity were enriched in the TBS extraction as compared to the isopropanol extraction: 2249 (35.7% of total) versus 469 (15.6%) total spectral count, respectively. In contrast, proteins annotated as having serine-type endopeptidase inhibitor activity were enriched by the isopropanol method: 697 (23.2% of total) for isopropanol compared with 422 for TBS (6.7% of total), and 403 for ethanol (4.8% of total). It is important to note that both the total number of proteins identified and their relative abundance (as measured by total spectral count) are not equal among the extraction methods. The TBS and ethanol extractions contained 3-fold more individual proteins than the isopropanol extraction: 908, 938, and 311 individual proteins were identified, respectively. Even though the TBS and ethanol extraction methods were able to purify a higher total number of proteins than the isopropanol method, the isopropanol method was substantially more efficient at enriching protease inhibitors and other lower abundance seed proteins (Figures 3 and 4). The isopropanol extraction method also revealed proteins associated with toxin activity at higher abundance (12.2% of total spectral count) as compared with the ethanol or TBS methods (2.6 and 1.5% of total spectral counts, respectively). These quantitative results suggest that the isopropanol extraction method is more efficient at extracting protease inhibitors and proteins annotated as having toxin activity than the TBS and ethanol extraction methods.

Differences and Similarities in Grass Pea Proteins Enriched by Three Different Protein Extraction Methods. To understand which proteins overlapped in each extraction, a Venn diagram was created using the top 25 proteins from each method using GOSlim terms (Figure 5). It is important to note that some proteins have multiple annotations and may appear several times in the same analysis. Four proteins overlapped from all three extractions: a protease inhibitor identified as the Bowman−Birk protease inhibitor, storage proteins glycinin and albumin, storage proteins glycinin and albumin, and proteins related to toxin activity. When looking at the proteins that were only extracted by TBS, the majority are storage proteins, mainly globulins. The isopropanol extraction was the only method that was able to enrich trypsin protease inhibitors. In addition, this method was efficient at extracting metal binding proteins and enzymes that are crucial for biochemical pathways; these comprised ∼37.8% of total proteins as compared with ∼25.5% of total proteins from the ethanol extraction [Table S1: Summarized Spectrum Data (all)].

In conclusion, these data suggest that the 50% isopropanol extraction method is the most useful method for proteomic analysis for enriching lower abundance proteins in grass pea seeds. This method is also very efficient at isolating protease inhibitors and proteins annotated as having toxic activity. The ethanol extraction method, in contrast, may be more useful for isolating enzymes associated with the synthesis of β-ODAP. Both extraction methods provide useful tools for breeders to detect and quantify expression levels of protease inhibitors and could be used to detect differences in expression levels in experimental lines toward release of new germplasm with lower levels of protease inhibitors.

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.9b04307.

Protein spectral data and GO assignments (XLSX)

Distribution of species for top BLAST hit for BLAST2GO analysis for 1376 proteins identified by peptide matches from three different extraction methods on grass pea mature seeds (PDF)

Distribution of GO annotation results using BLAST2-GO for 1376 proteins identified by peptide matches from three different extraction methods on grass pea mature seeds (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hari.b.krishnan@usda.gov.

ORCID

Hari B. Krishnan: 0000-0001-6437-3672

Author Contributions

⊥C.M. and Q.X. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the USDA-Agricultural Research Service and China Postdoctoral Science Foundation (2016M590975) and Postdoctoral Science Foundation of Shaanxi province (2016BSHEDZZ119), P.R. China. Mention of a trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

■ ABBREVIATIONS

TBS, Tris-buffered saline; β-ODAP, β-*N*-oxalyl-L-α,β-diaminopropionic acid; ISO, isopropanol; EtOH, ethanol

■ REFERENCES

(1) Gurung, A. M.; Pang, E. C. Lathyrus. *Wild Crop Relatives: Genomic and Breeding Resources*; Springer, 2011; pp 117−126.

(2) Campbell, C. G. Grass pea, *Lathyrus sativus* L. *Biodivers. Int. J*: **1997**; Vol. *18*.

(3) Tan, R.-Y.; Xing, G. Y.; Zhou, G. M.; Li, F. M.; Hu, W. T.; Lambein, F.; Xiong, J. L.; Zhang, S. X.; Kong, H. Y.; Zhu, H. Plant toxin β-ODAP activates integrin β1 and focal adhesion: A critical pathway to cause neurolathyrism. *Sci. Rep.* **2017**, *7*, 40677.

(4) Xu, Q.; Liu, F.; Chen, P.; Jez, J.; Krishnan, H. β-diaminopropionic acid (β-ODAP) content in *Lathyrus sativus*: the integration of nitrogen and sulfur metabolism through β-cyanoalanine synthase. *Int. J. Mol. Sci.* **2017**, *18*, 526.

(5) Yan, Z.-Y.; Spencer, P. S.; Li, Z.-X.; Liang, Y.-M.; Wang, Y.-F.; Wang, C.-Y.; Li, F.-M. *Lathyrus sativus* (grass pea) and its neurotoxin ODAP. *Phytochemistry* **2006**, *67*, 107−121.

(6) Almeida, N. F.; Rubiales, D.; Patto, M. C. V. Grass Pea. *Grain Legumes*; Springer, 2015; pp 251−265.

(7) Hanbury, C. D.; White, C. L.; Mullan, B. P.; Siddique, K. H. M. A review of the potential of *Lathyrus sativus* L. and *L. cicera* grain for use as animal feed. *Anim. Feed Sci. Technol.* **2000**, *87*, 1−27.

(8) Kumar, S.; Bejiga, G.; Ahmed, S.; Nakkoul, H.; Sarker, A. Genetic improvement of grass pea for low neurotoxin (β-ODAP) content. *Food Chem. Toxicol.* **2011**, *49*, 589−600.

(9) Fikre, A.; Van Moorhem, M.; Ahmed, S.; Lambein, F.; Gheysen, G. Studies on neurolathyrism in Ethiopia: dietary habits, perception of risks and prevention. *Food Chem. Toxicol.* **2011**, *49*, 678−684.

(10) Lambein, F.; Travella, S.; Kuo, Y.-H.; Van Montagu, M.; Heijde, M. Grass pea (*Lathyrus sativus* L.): orphan crop, nutraceutical or just plain food? *Planta* **2019**, *250*, 821−838.

(11) Kuo, Y.-H.; Ikegami, F.; Lambein, F. Metabolic routes of β-(isoxazolin-5-on-2-yl)-L-alanine (BIA), the precursor of the neurotoxin ODAP (β-n-oxalyl-L-α, β,-diaminopropionic acid), in different legume seedlings. *Phytochemistry* **1998**, *49*, 43−48.

(12) Spencer, P.; Ludolph, A.; Dwivedi, M. P.; Roy, D.; Hugon, J.; Schaumburg, H. H. Lathyrism: evidence for role of the neuro-excitatory aminoacid BOAA. *Lancet* **1986**, *328*, 1066−1067.

(13) Jiao, C.-J.; Jiang, J.-L.; Ke, L.-M.; Cheng, W.; Li, F.-M.; Li, Z.-X.; Wang, C.-Y. Factors affecting β-ODAP content in *Lathyrus sativus* and their possible physiological mechanisms. *Food Chem. Toxicol.* **2011**, *49*, 543−549.

(14) Jiao, C.-J.; Xu, Q.-L.; Wang, C.-Y.; Li, F.-M.; Li, Z.-X.; Wang, Y.-F. Accumulation pattern of toxin β-ODAP during lifespan and effect of nutrient elements on β-ODAP content in *Lathyrus sativus* seedlings. *J. Agric. Sci.* **2006**, *144*, 369−375.

(15) Getahun, H.; Mekonnen, A.; TekleHaimanot, R.; Lambein, F. Epidemic of neurolathyrism in Ethiopia. *Lancet* **1999**, *354*, 306−307.

(16) Ravindranath, V. Neurolathyrism: mitochondrial dysfunction in excitotoxicity mediated by L-β-oxalyl aminoalanine. *Neurochem. Int.* **2002**, *40*, 505−509.

(17) Haimanot, R. T.; Kidane, Y.; Wuhib, E.; Kalissa, A.; Alemu, T.; Zein, Z. A.; Spencer, P. S. Lathyrism in rural northwestern Ethiopia: a highly prevalent neurotoxic disorder. *Int. J. Epidemiol.* **1990**, *19*, 664−672.

(18) Hailu, D.; Abera, S.; Teka, T. A.; Box, P.; Jimma, E. Effects of Processing on Nutritional Composition and Anti-Nutritional Factors of Grass pea (*Lathyrus Sativus* L): A review. *Food Sci. Qual. Manag.* **2015**, *36*, 61−71.

(19) Lambein, F.; Khan, J. K.; Kuo, Y.-H.; Campbell, C. G.; Briggs, C. J. Toxins in the seedlings of some varieties of grass pea (*Lathyrus sativus*). *Nat. Toxins* **1993**, *1*, 246−249.

(20) Kebede, B.; Urga, K.; Nigatu, A. Effect of processing methods on the trypsin inhibitor, tannins, phytic acid and ODAP contents of grass pea seeds. *Ethiop. J. Health Dev.* **1995**, *9*, 97−103.

(21) Xu, Q.; Qu, J.; Song, B.; Liu, F.; Chen, P.; Krishnan, H. B. *Lathyrus sativus* originating from different geographical regions reveals striking differences in Kunitz and Bowman-Birk inhibitor activities. *J. Agric. Food Chem.* **2019**, *67*, 8119−8129.

(22) Laskowski, M., Jr; Kato, I. Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **1980**, *49*, 593−626.

(23) Krishnan, H. B.; Oehrle, N. W.; Natarajan, S. S. A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: a case study using *Glycine max*. *Proteomics* **2009**, *9*, 3174−3188.

(24) Byers, M.; Miflin, B. J.; Smith, S. J. A quantitative comparison of the extraction of protein fractions from wheat grain by different solvents, and of the polypeptide and amino acid composition of the alcohol-soluble proteins. *J. Sci. Food Agric.* **1983**, *34*, 447−462.

(25) Natarajan, S. S.; Krishnan, H. B.; Lakshman, S.; Garrett, W. M. An efficient extraction method to enhance analysis of low abundant proteins from soybean seed. *Anal. Biochem.* **2009**, *394*, 259−268.

(26) Krishnan, H. B.; Natarajan, S. S.; Oehrle, N. W.; Garrett, W. M.; Darwish, O. Proteomic analysis of pigeonpea (*Cajanus cajan*) seeds reveals the accumulation of numerous stress-related proteins. *J. Agric. Food Chem.* **2017**, *65*, 4572−4581.

(27) Talbot, N. C.; Shannon, A. E.; Phillips, C. E.; Garrett, W. M. Derivation and characterization of a pig embryonic-stem-cell−derived exocrine pancreatic cell line. *Pancreas* **2017**, *46*, 789−800.

(28) He, L.; Diedrich, J.; Chu, Y.-Y.; Yates, J. R., III Extracting accurate precursor information for tandem mass spectra by RawConverter. *Anal. Chem.* **2015**, *87*, 11361−11367.

(29) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications

made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383−5392.
(30) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646−4658.
(31) Conesa, A.; Gotz, S.; Garcia-Gomez, J. M.; Terol, J.; Talon, M.; Robles, M. a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674−3676.
(32) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed storage proteins: structures and biosynthesis. *Plant Cell* **1995**, *7*, 945−956.
(33) Marmagne, A.; Rouet, M.-A.; Ferro, M.; Rolland, N.; Alcon, C.; Joyard, J.; Garin, J.; Barbier-Brygoo, H.; Ephritikhine, G. Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. *Mol. Cell. Proteomics* **2004**, *3*, 675−691.
(34) Ferro, M.; Salvi, D.; Brugière, S.; Miras, S.; Kowalski, S.; Louwagie, M.; Garin, J.; Joyard, J.; Rolland, N. Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Mol. Cell. Proteomics* **2003**, *2*, 325−345.
(35) Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. M. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal. Biochem.* **2005**, *342*, 214−220.
(36) Lei, Z.; Elmer, A. M.; Watson, B. S.; Dixon, R. A.; Mendes, P. J.; Sumner, L. W. A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*. *Mol. Cell. Proteomics* **2005**, *4*, 1812−1825.
(37) Krishnan, B. B. A simple and rapid method to isolate low molecular weight proteinase inhibitors from soybean. *Korean J. Crop Sci.* **2004**, *49*, 342−348.
(38) Duval, M.; Job, C.; Alban, C.; Douce, R.; Job, D. Developmental patterns of free and protein-bound biotin during maturation and germination of seeds of *Pisum sativum*: characterization of a novel seed-specific biotinylated protein. *Biochem. J.* **1994**, *299*, 141−150.
(39) Wehmeyer, N.; Vierling, E. The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiol.* **2000**, *122*, 1099−1108.
(40) Boudet, J.; Buitink, J.; Hoekstra, F. A.; Rogniaux, H.; Larré, C.; Satour, P.; Leprince, O. Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. *Plant Physiol.* **2006**, *140*, 1418−1436.
(41) Haslekås, C.; Viken, M. K.; Grini, P. E.; Nygaard, V.; Nordgard, S. H.; Meza, T. J.; Aalen, R. B. Seed 1-cysteine peroxiredoxin antioxidants are not involved in dormancy, but contribute to inhibition of germination during stress. *Plant Physiol.* **2003**, *133*, 1148−1157.
(42) Ball, R. O.; Courtney-Martin, G.; Pencharz, P. B. The in vivo sparing of methionine by cysteine in sulfur amino acid requirements in animal models and adult humans. *J. Nutr.* **2006**, *136*, 1682S−1693S.
(43) Giannopolitis, C. N.; Ries, S. K. Superoxide dismutases: I. Occurrence in higher plants. *Plant Physiol.* **1977**, *59*, 309−314.
(44) Laporte, D.; Olate, E.; Salinas, P.; Salazar, M.; Jordana, X.; Holuigue, L. Glutaredoxin GRXS13 plays a key role in protection against photooxidative stress in Arabidopsis. *J. Exp. Bot.* **2011**, *63*, 503−515.
(45) Calvo, E. S.; Wurtle, E. S.; Shoemaker, R. C. Cloning, mapping, and analyses of expression of the Em-like gene family in soybean [*Glycine max* (L). Merr.]. *Theor. Appl. Genet.* **1997**, *94*, 957−967.
(46) Schaefer, H.; Hechenleitner, P.; Santos-Guerra, A.; de Sequeira, M. M.; Pennington, R. T.; Kenicer, G.; Kenicer, G. Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evol. Biol.* **2012**, *12*, 250.
(47) Smýkal, P.; Kenicer, G.; Flavell, A. J.; Corander, J.; Kosterin, O.; Redden, R. J.; Ford, R.; Coyne, C. J.; Maxted, N.; Ambrose, M. J. Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. *J. Plant Genet. Resour.* **2011**, *9*, 4−18.
(48) Smýkal, P.; Coyne, C. J.; Ambrose, M. J.; Maxted, N.; Schaefer, H.; Blair, M. W.; Berger, J.; Greene, S. L.; Nelson, M. N.; Besharat, N. crops phylogeny and genetic diversity for science and breeding. *CRC Crit. Rev. Plant Sci.* **2015**, *34*, 43−104.

DOI: 10.1021/acs.jafc.9b04307
J. Agric. Food Chem. 2019, 67, 10296−10305