

Protein Profiling of Non-model Plant *Cuminum cyminum* by Gel-Based Proteomic Approach

Uzma Zaman,^{a,b,c,*} Henning Urlaub^{a,b} and Atiya Abbasi^d

ABSTRACT:

Introduction – Cumin (*Cuminum cyminum*), a popular spice has been widely used in traditional medicine to cure various ailments. Despite the existence of scientific literature about its pharmacological properties, no successful proteome profiling has yet been attempted.

Objective – To optimise extraction of cumin proteins and analyse its profile by shotgun proteomics, using one-dimensional electrophoresis coupled with nano-ESI-LC-MS/MS.

Methodology – As a first step, we have compared three extraction protocols for total proteins extraction from cumin. Extracted proteins were separated on one-dimensional gel and analysed by state-of-the-art linear ion trap (LTQ)-Orbitrap Velose and Q Exactive HF mass spectrometer.

Results – Evaluation of extraction method revealed significant differences in protein yield and proteome composition between the three extracts. LC-MS/MS allowed identification of several proteins with functional significance in various biological processes.

Conclusion – This study provides identification of a large number of proteins and offers a molecular basis for future research on potential pharmacologically active cumin proteins. Copyright © 2017 John Wiley & Sons, Ltd.

Additional Supporting Information may be found online in the supporting information tab for this article.

Keywords: *Cuminum cyminum*; cumin; liquid chromatography-high resolution mass spectrometry; one-dimensional gel electrophoresis; plant proteomics; protein extraction

Introduction

Cumin (*Cuminum cyminum*), seeds of herbaceous plant belonging to family Apiaceae, is indigenous to the Mediterranean region and has been popularly used as a spice. Studies report that cumin exhibits a broad range of pharmacological properties, including antibacterial (Gachkar et al., 2007), antifungal (Hajlaoui et al., 2010), antioxidant (El-Ghorab et al., 2010), hypoglycemic (Jagtap and Patil, 2010) and anti-carcinogenic (Ekta Prakash, 2014). Although cumin seeds contain essential oils, proteins, carbohydrates, vitamins and minerals, only the chemical constituents and pharmacological aspects of the essential oils have been thoroughly investigated (Mnif and Aifa, 2015) and no valuable work has been reported so far for the other components including proteins. Nomura et al. (1999) first reported starch degrading α and β -amylase activity in cumin seeds by viscosity measurements test, reducing sugar assays and TLC-analysis. We have earlier reported isolation and purification of a non-specific lipid transfer protein 1 (nsLTP1) from cumin aqueous extract. nsLTP1 was purified by using classical protein chemistry techniques for elucidation of its primary structure (Zaman and Abbasi, 2009). Then, Masoumi et al. (2012) analysed the seed storage protein pattern by SDS-PAGE electrophoresis to evaluate genetic diversity and classification of plants from the family Apiaceae. Recently, proteomic screening by mass spectrometry (MS) has been applied to investigate the presence of peanut-specific allergen proteins in cumin samples from different sources (Garber et al., 2016). The study was conducted to establish a method to determine the presence of undeclared food allergens and the investigation focused on the identification of allergen proteins only.

The present work is aimed at comprehensive proteomic analysis of cumin by applying a gel-based approach coupled to nanoscale liquid chromatography with tandem mass spectrometry (nano-LC-MS/MS). Although the field of plant proteomics has grown in recent years, it still lags behind human and animal proteomics. The major technical obstacle in plant proteomics is the efficient extraction of total proteins, separation of protein complex and database processing for functional analysis in order to interpret the biological significance of identified proteins.

The most critical step is the selection of the protein extraction method to allow the purification of as many proteins while minimising co-purification of non-protein contaminants. In the present study we have compared three different extraction methods for preparation of protein samples suitable for gel

* Correspondence to: Uzma Zaman, Dow Research Institute of Biotechnology and Biomedical Sciences, Dow University of Health Sciences, Gulzar-e-Hijri, Suparco Road, KDA Scheme-33, Karachi, Pakistan. Email: uzma.zaman@duhs.edu.pk

^a Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077, Göttingen, Germany

^b Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Centre, Robert-Koch-Strasse 40, D-37075, Göttingen, Germany

^c Dow Research Institute of Biotechnology and Biomedical Sciences, Dow University of Health Sciences, Gulzar-e-Hijri, Suparco Road, KDA Scheme -33, Karachi, Pakistan

^d International Centre for Chemical and Biochemical Sciences (ICCBS), HEJ Research Institute of Chemistry, University of Karachi, Karachi -75270, Pakistan

electrophoresis. The first method utilises sodium acetate extraction at slightly acidic conditions (pH 5.5) followed by ammonium sulphate (SA/AS) precipitation. This method has been successfully applied and reported for the purification of proteins from cumin (Zaman and Abbasi, 2009). The second method uses Tris saturated phenol solution for protein extraction followed by precipitation with methanolic ammonium acetate. This method is routinely applied in plant proteomics and is quite popular due to its ability to efficiently remove non-protein contaminants during extraction (Faurobert *et al.*, 2007) although it involves many washing steps which are quite laborious and time consuming. For the third method, owing to the importance of whole proteome identification from plant and animal sources a number of different ready-mix-reagents have also been developed which allow robust extraction of total proteins. We have therefore selected Readyprep total protein extraction kit from Bio-Rad (Hercules, CA, USA) and evaluate its suitability for the extraction of cumin proteins.

Gel-based approach particularly, two-dimensional gel electrophoresis (2DE) has been widely applied in plant proteome research as a separation technique. However, the method is very time consuming and has reproducibility issues along with the loss of low abundant proteins (Gygi *et al.*, 2000). In addition the 2DE method has limited capability to resolve hydrophobic protein, highly acidic/basic proteins and high molecular proteins (Ong and Pandey, 2001). To overcome these challenges, we preferred to use one-dimensional gel electrophoresis (1DE) for the current studies. After separation of total extracted protein on 1DE gel, the entire gel lane was cut into slices, subjected to the proteolytic digestion using trypsin and the resulting peptides were separated using the nano-LC system prior to MS/MS analysis. The use of a high-resolution mass spectrometer such as the hybrid linear ion trap (LTQ)-Orbitrap Velose allowed the identification of several proteins. Analysis of the protein samples using modern state-of-the-art Q Exactive HF mass spectrometer, which allows fast speed as well as high mass resolution and high sensitivity, improved the identification of proteins by several fold. Acquired MS data were processed using MaxQuant (Cox and Mann, 2008) and identified peak lists were submitted to the Andromeda search engine (Cox *et al.*, 2011) (www.maxquant.org).

A major challenge in plant proteomics, as MS-based proteomic analysis depends on high-quality sequence databases, is the absence of well-annotated and complete genome sequences. Recently, plant proteomic literature has increased considerably to include most model plants and cash crops (e.g. Arabidopsis and rice) while cumin as a non-model plant, suffers with lack of genomic and proteomic data. Therefore, proteins were identified by homology with other plant species. To validate the identified proteins list, the intensity-based absolute quantification (iBAQ) option was enabled in the MaxQuant software. The iBAQ algorithm allowed estimation of protein abundances by dividing the sum of all the peptide intensities with the number of observable peptides of a protein (Schwanhaussner *et al.*, 2011). The resulting iBAQ intensities provide an accurate determination of the relative abundance of all proteins identified in a sample. Therefore, we have narrowed down the identified list of proteins by evaluating the iBAQ intensities. Furthermore, biological significance of identified proteins was done by PANTHER (Protein Analysis Through Evolutionary Relationships) (Mi *et al.*, 2013), which is a comprehensive database. It was used to analyse gene ontology and protein family of identified proteins. To our knowledge, this is the first study aimed at full characterisation of cumin proteome. Our approach with the use of high resolution and high mass accuracy instrument and with iBAQ

intensities provides high-confidence identifications of peptides. This approach can be applied to other plant species as well.

Materials and methods

Chemicals and reagents

All chemicals, reagents, and organic solvents were purchased from Sigma Aldrich (St Louis, MO, USA). Readyprep protein extraction kit (catalogue number 163–2086) was purchased from Bio-Rad. Ammonium bicarbonate, dithiothreitol (DTT) and iodoacetamide were purchased from Sigma-Aldrich. MS grade trypsin from Serva (Heidelberg, Germany), 4–15% NuPAGE Novex Bis-Tris gel from Invitrogen (Thermo Fisher Scientific, Waltham, MA, USA). Precision Plus Protein™ Unstained Standards (10–250 kDa) from Bio-Rad, HPLC grade solvents such as Lichrosolv Water and Lichrosolv Acetonitrile were purchased from Merck (Darmstadt, Germany). Formic acid and Coomassie Brilliant Blue were from Fluka (Buchs, Germany), and Amresco Inc. (Solon, OH, USA) respectively.

Protein extraction

Cuminum cyminum (Cumin) was purchased from a local supplier. Finely grinded cumin seeds were subjected to total protein extraction by three methods described later in detail.

Sodium acetate/ammonium sulphate (SA/AS) extraction

One gram of powdered *Cuminum cyminum* was soaked in 10 mL of buffer A (25 mM sodium acetate pH 5.5, 4 mM EDTA, 0.01% NaN₃) overnight at 4°C with continuous stirring. Supernatant was collected by centrifugation at 6000×g (4°C) for 30 min. Proteins were precipitated using ammonium sulphate (70%). After overnight incubation at 4°C, precipitates were collected by centrifugation at 10000×g (4°C) for 20 min. Precipitates were re-suspended in buffer A for protein estimation and further analysis.

Phenol/ammonium acetate (Ph/AA) extraction

One gram of powdered *Cuminum cyminum* was soaked in 10 mL of buffer B (50 mM Tris-HCl pH 7.5, 15 mM EDTA, 5 mM KCl, 5 mM DTT) at 4°C with continuous stirring overnight. After centrifugation at 10000×g (4°C) for 30 min, the supernatant was collected and equal volumes of Tris-HCl saturated phenol was added. The upper phenol layer was collected by centrifugation at 10000×g (4°C) for 15 min. The proteins were extracted by precipitation with five volumes (v/v) of 100 mM ammonium acetate in methanol. After overnight incubation at –20°C, protein precipitates were collected by centrifugation at 10000×g (4°C) for 5 min. The pellet was washed three times with 100 mM ammonium acetate in methanol and once with acetone (100%) and then air dried.

Bio-rad Readyprep protein extraction kit (kit)

Readyprep protein extraction kit from Bio-Rad was used according to the manufacture instructions. Readyprep 2D sample buffer and Readyprep TBP reducing agent were reconstituted in proteomic grade water. Then, 50 µL of Readyprep TBP reducing agent was mixed with 5 mL Readyprep 2D sample buffer. Next, 1 mL of buffer was added to 500 mg of fine cumin powder, sonicated briefly and centrifuged at 10000×g (4°C) for 30 min. Supernatant was used for protein quantification and gel electrophoresis.

Protein quantification

The amount of protein recovered from cumin using the three different extraction methods was estimated by the Bradford method using the protein assay kit (Thermo Fisher Scientific) (Bradford, 1976). A standard curve was prepared by known concentration of bovine serum albumin (BSA 1 mg/mL).

One-dimensional gel electrophoresis (1DE)

From each extract 100 µg of proteins were resuspended in NuPAGE-LDS sample buffer (4×) and heated for 10 min at 70°C. The heated sample was vortexed and after a short spin was loaded on pre-casted NuPAGE Novex Bis-Tris gel (4–15%). Gel was run with MOPS-SDS running buffer at constant voltage (200 V) for 60 min under reducing condition and later stained with colloidal Coomassie Brilliant Blue R-250 overnight. De-staining was done by water for two to three hours (Schmidt and Urlaub, 2009).

In-gel digestion

Sample lanes of gel were cut into 23 gel slices using in-house built gel cutter (Schmidt and Urlaub, 2009). All the gel slices were further cut into small pieces and washed first with water then with 100% acetonitrile at 37°C with shaking. In-gel protein reduction and alkylation was carried out with 10 mM DTT reagent at 56°C for 60 min and 50 mM iodoacetamide in the dark at room temperature for 20 min, respectively. Excess reagent was washed with water at 37°C for 15 min followed by dehydration using 50% acetonitrile for 15 min. Gel pieces were allowed to air-dry for 5 to 10 min. The gel pieces were incubated in trypsin solution initially at 4°C for 20 min to absorb the trypsin and subsequently at 37°C overnight. Peptides were extracted by incubating the gel pieces in 5% formic acid and were dried in a SpeedVac. For LC-MS analysis dried peptides were dissolved in 20 µL of buffer (5% acetonitrile, 1% formic acid solution) and sonicated for 1 min.

Nano-LC-MS/MS analysis

Peptide digest was applied onto a nanoflow liquid chromatograph system (Agilent 1100 series, Agilent technologies, Boblingen, Germany) coupled to an LTQ-Orbitrap Velose mass spectrometer using a Flex Ion nano Spray source (Thermo Scientific, Schwerte, Germany). Peptides were separated on a self-packed analytical reverse phase-C18 column (150 mm length, 0.075 mm inner diameter, ReproSil-Pur C₁₈-AQ 3 µm resin, Dr Maisch GmbH) using a linear gradient of 3 to 36% buffer B (80% acetonitrile and 0.1% formic acid) for 37 min. A Top 15 method was used with the collision-induced dissociation (CID) mode of fragmentation. Full MS scans were acquired over the *m/z* 350–1600 range with resolution 30000 at *m/z* 400 with the Orbitrap mass analyser. The 15 most intense peaks with charge state ≥ 2 were selected for fragmentation with normalised collision energy of 35%, activation time of 10 ms, and one microscan. Other parameters include, ion selection threshold –2000 counts, ion accumulation times –500 ms for full scans and 100 ms for CID (Atanassov and Urlaub, 2013).

Further in-depth protein profiling of cumin extract was done by high speed Q Exactive HF (Thermo Fisher Scientific, Bremen, Germany). Digested peptides were applied on UltiMate 3000 RSLC-nano HPLC system (Thermo Fisher Scientific) coupled online to the Q Exactive HF. Peptides were separated on analytical

column (30 cm long, 75 µm inner diameter) packed in-house with ReproSil-Pur C18 AQ 1.9 µm reversed phase resin (Dr Maisch GmbH). A linear gradient of 5 to 45% buffer B (80% acetonitrile and 0.1% formic acid) was used over 88 min at a flow rate of 300 nL/min. The column temperature was maintained at 50°C. The MS data was acquired in mass range from 350 to 1600 Da at a resolution of 60000 at *m/z* 200. Top 30 ion were selected for fragmentation in the high energy collision-induced dissociation (HCD) mode at a resolution of 15000 at *m/z* 200 with maximum IT of 50 ms. The precursors were isolated with a window of 2.0 Da and HCD fragmentation was performed with the AGC (automatic gain control) target fill value of 1e⁵ ions (Pan *et al.*, 2017).

Data analysis

All raw data files from Xcalibur software were directly analysed by MaxQuant (Cox and Mann, 2008) (version 1.3.0.5) using the Andromeda Search engine (Cox *et al.*, 2011) against the Viridiplantae database from UniProt. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino *et al.*, 2014) partner repository with the dataset identifier PXD006257. Carbamidomethylation of cysteine was specified as fixed modification while oxidation of methionine and N-terminal acetylation were considered as variable modifications. Minimum peptide length was set to six amino acids. Minimum number of unique peptides was set to one. The maximum false discovery rate (FDR) for both peptide and protein identifications was set to 1%, calculated by employing a reverse database strategy. Proteins identified as “reverse” were discarded from the list of identified proteins. Strict specificity for trypsin with no proline restriction and maximum two mis-cleavages were allowed. The initial precursor mass tolerance was 6 ppm, and for the fragment masses, it was up to 20 ppm. The iBAQ (Schwanhauser *et al.*, 2011) in MaxQuant was performed on the identified peptides to quantify protein abundance.

Bioinformatics analysis

Due to lack of protein sequences and gene sequence information from cumin, protein identification was based upon different species in Viridiplantae database. To investigate the functional properties, all the identified proteins were blasted in the entire UniProtKB database, filtering the result by *Arabidopsis thaliana* protein database (<http://www.arabidopsis.org/>) to obtain annotated protein entries. Results with the highest score and lowest E value were considered as relevant for each identified protein. This list of proteins was then submitted to a gene ontology (GO) analysis using Panther 9.0 (www.pantherdb.org/) (Mi *et al.*, 2013). PANTHER GO Classification tool categorised proteins according to their Biological Processes, Molecular Functions, and Cellular Components.

Results and discussion

In the present study, we have proposed a gel-based approach for comprehensive proteomic characterisation of cumin. Despite the potential medicinal importance, cumin has not been extensively studied in the proteomic field. In this work, we tried different extraction methods to achieve best results. Extraction of proteins from plant family, such as Apiaceae, is a big challenge due to low protein content. For this reason, we set out to evaluate three different protein extraction methods suitable for analysis and provide a

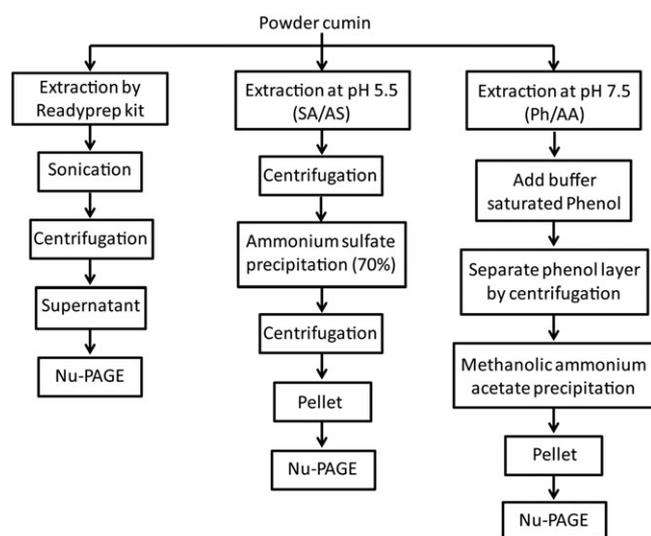


Figure 1. Schematic outline of three extraction methods used in this study.

proteomic platform that could be used for a comprehensive qualitative description of cumin proteome. A summary of these methods is shown in Fig. 1. The experiments were carried out in duplicate and for each sample three technical replicates were performed to assess the analytical reproducibility.

In the SA/AS extraction method, cumin proteins were extracted at pH 5.5. The key point of this method was to isolate proteins under slightly acidic conditions as solubility is a prerequisite for the protein yield from plants. Generally, very high and low pH-values increase protein solubility but possibly at the expense of structural changes of the proteins. Previously we have reported a similar buffer system (pH 5.5) for purification of nSLTP1 from cumin (Zaman and Abbasi, 2009). Thus, for a more comprehensive proteomic study we have utilised slightly acidic soluble fraction in order to provide a better understanding of the characteristics and functionality of cumin proteins. The extraction was followed by ammonium sulphate precipitation to isolate and clean-up protein extracts from interfering material. In plants, polysaccharides, phenolic compounds and secondary metabolites severely interfere in proteomic studies (Wang *et al.*, 2008). Extracted proteins were applied on 1DE. Figure 2 shows the gel image. This method showed a high degree of smearing and non-migrating materials in the gel indicating incomplete removal of interfering material during the extraction procedure. The protein bands were also not well resolved.

The selection of an ideal extraction method for reproducible capture of proteins by minimising protein degradation and non-protein contaminants is a very crucial step in plant proteomics. The phenol-based method is very popular for total protein extraction from plant tissues that effectively remove nucleic acids, ions and polysaccharides (Wang *et al.*, 2008) from crude extract. In this method, proteins extracted with Tris-HCl buffer (pH 7.5), are further extracted with Tris-HCl buffer saturated phenol. Proteins dissolved in the phenolic phase were then precipitated with chilled ammonium acetate/methanol solution (Yan Zehn, 2011). The analysis of the gel image (Fig. 2) of purified proteins showed slightly distorted bands with less resolution. The band patterns of SA/AS and Ph/AA methods were almost similar to each other, however, comparatively less smearing and no accumulated materials in the gel was observed with the Ph/AA method (Fig. 2).

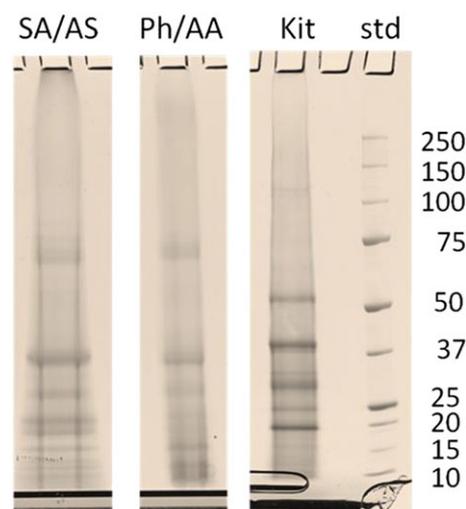


Figure 2. One-dimensional electrophoresis of cumin proteins from three extraction methods. From the left, lane labelled with SA/AS, Ph/AA and Kit shows results of three extraction methods, sodium acetate/ammonium sulphate, phenol/ammonium acetate and Readyprep protein extraction kit, respectively. Lane labelled as Std contain a molecular marker with the molecular weights of the reference proteins marked. [Colour figure can be viewed at wileyonlinelibrary.com]

Despite recent developments in protein extraction methods for comprehensive proteomic analysis of notoriously challenging samples, the need for a fast, uncomplicated, and robust method is still required. The Readyprep protein extraction kit (Kit) provides a one-step protocol for extraction of total protein with high reproducibility. The Kit contains urea, thiourea and zwitterionic detergent ASB-14 in a buffer that extract total protein suitable for gel analysis. Furthermore, tributylphosphine, reducing agents enhance solubilisation of proteins. The analysis of protein yield showed that the Kit method gave significantly greater protein yield than the other two methods (Table 1).

Evaluation of extraction method on gels revealed clear differences in terms of both the pattern and resolution of protein bands between the three methods (Fig. 2). Visual comparison of the gel shows well resolved and more intense bands with the Kit method indicating the generation of a better quality sample free of interfering material.

Further evaluation of extraction methods was done by LC-MS/MS analysis of extracted protein/peptides followed by evaluation with MaxQuant software (Cox and Mann, 2008). A major challenge for large-scale proteomic study from plant sample is the absence and limited availability of protein sequences in databases due to lack of whole genome sequence. Cumin, an unsequenced species, belongs to non-model plant species. Thus, characterisation

Table 1. Protein yields of cumin from three extraction methods calculated by Bradford assay

Extraction method	Protein yield ($\mu\text{g}/\mu\text{L}$)
Sodium acetate/ammonium sulphate (SA/AS)	0.94 ± 0.26
Phenol/ammonium acetate (Ph/AA)	1.34 ± 0.18
Readyprep protein extraction kit (Kit)	2.08 ± 0.33

of cumin proteomic data was carried out principally by homology with several plant species. For this reason we searched data against whole viridiplantae database without restriction to any specific specie resulting in the identification of a large number of peptides with sequence homology to proteins belonging to several species. To improve the validity of the obtained results, data was manually checked and filtered under strict criteria. First, database redundancy was avoided by filtering out proteins that appeared more than once. Moreover, we also considered the total number of identified peptides, number of unique peptides and sequence coverage. Finally, the iBAQ intensities and proteins having large score were accepted.

A total of 657 proteins were identified by merging the data from all three methods. The complete lists of all identified proteins for each experiment are given in the Supporting Information Table S1. Venn diagrams in Fig. 3 illustrate the overlap of protein identification among the three methods. Among these, out of 479 proteins identified from the Kit method, 392 proteins were unique, 94 and 47 proteins appeared unique among the 219 and 132 protein identified from SA/AS and Ph/AA extraction methods, respectively. Among all identified proteins 37 were common to all three methods. Thus, the Kit method contributes 72% of total proteins identified.

To explore identified proteins in terms of their functional annotation and to investigate the difference in specificity of purification of proteins between the three extraction methods, PANTHER analysis (Mi *et al.*, 2013) was performed. Since the protein identification was accomplished by sequence homology with several different plant species, identified proteins were blasted against the *Arabidopsis thaliana* protein database. Figure 4 displays the functional classification of the identified proteins from all three extraction methods as a gene ontology pie chart. The identified proteins are grouped into three main categories such as molecular function, biological process and cellular components whereas the bar diagram shows the comparison between gene ontology

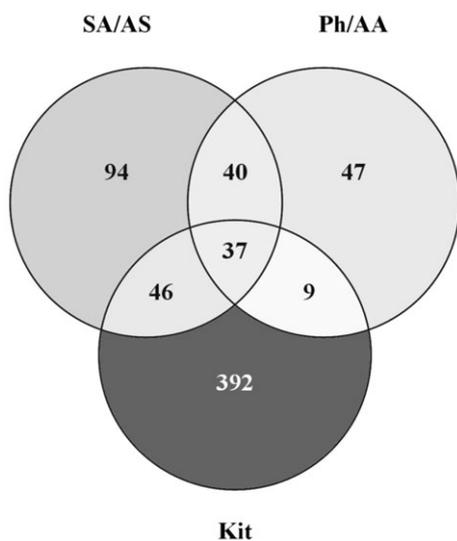


Figure 3. Venn diagram showing the distribution of proteins identified from cumin by using three extraction methods. SA/AS (sodium acetate/ammonium sulphate), Ph/AA (phenol/ammonium acetate) and Kit (Readyprep protein extraction kit). Number shared by two circles indicates proteins common in two methods. The number shared by all three circles represent the number of proteins common in all three methods.

proteins identifies from each extraction method (Fig. 4). Concerning molecular functions and biological processes, very minor differences were observed. These differences include the presence of eight transfer carrier proteins, which are exclusively identified by the Kit method (Supporting Information, Fig. S1). These include especially aquaporin, ADP/ATP carrier proteins and importin. Aquaporins localised in cell junction, were identified with good iBAQ intensities.

Interestingly with the SA/AS method, the percentage of ligases, kinases, hydrolases, proteases were high compared to the other two methods. Similarly, percentage of proteins that function as signalling molecules, transcription factors and receptor were also high whereas calcium binding proteins and carrier proteins were not identified by the SA/AS method. These results confirmed that the extraction methods are specific for proteins related to particular process.

As the number of identified proteins was less from the Ph/AA method, low abundant protein involved in translation regulation and signalling molecules and enzymes such as ligases were not observed. Although the Ph/AA method maximises the elimination of interfering material, it is quite time consuming and exhibited less number of protein bands and less number of identified proteins. Less number of proteins could be explained by lower detection of less abundant proteins.

Based on the gel electrophoresis and LC-MS/MS results obtained, we conclude that the Kit method with its robustness and simplicity provides high quality protein samples. Identification of comparatively high number of proteins pave the way to further high-throughput proteomic analyses of the cumin by the use of a high-resolution mass spectrometer, like Q-Exactive HF (Thermo Fisher Scientific, Bremen, Germany). High field Orbitrap not only allowed the identification in shorter time but also increased the number of identified proteins for comprehensive proteomic characterisation of cumin species.

The whole proteomic platform permitted the identification of 1321 proteins with a FDR of 1% (Supporting Information, Table S4). Among these, 933 protein assignments were based on single unique peptide identifications. The number of unique peptide per protein was found to be very low because protein assignment was performed by homology against databases of whole green plant species. As expected, high sensitivity and fast speed high field orbitrap provided additional identification of proteins present in low amount leading to identification of a very large number of proteins compared to the Velose data.

Further functional classification of proteins was done after blast and filtering the result by *Arabidopsis thaliana* protein database (<http://www.arabidopsis.org/>) to obtain annotated protein entries. The proteins for which no known function could be assigned accounted for 42% of all identified proteins. However, proteins with known function were further classified by PANTHER GO annotation tool into, molecular function, biological process and cellular component (Fig. 5). With regard to the molecular function PANTHER analysis showed that proteins were mostly associated with catalytic (46.6%), binding (19.0%) and structural molecular activity (9.5%) (Fig. 5). This study also revealed the association of proteins in different biological processes, including metabolic processes (63.0%), cellular processes (19.2%), localisation (12.2%), response to stimulus (7.10%), biological regulation (6.9%), among others (Fig. 5; Supporting Information, Table S5).

Cellular component analysis indicated that major fraction of the identified proteins belongs to the cytoplasm (20%), intracellular

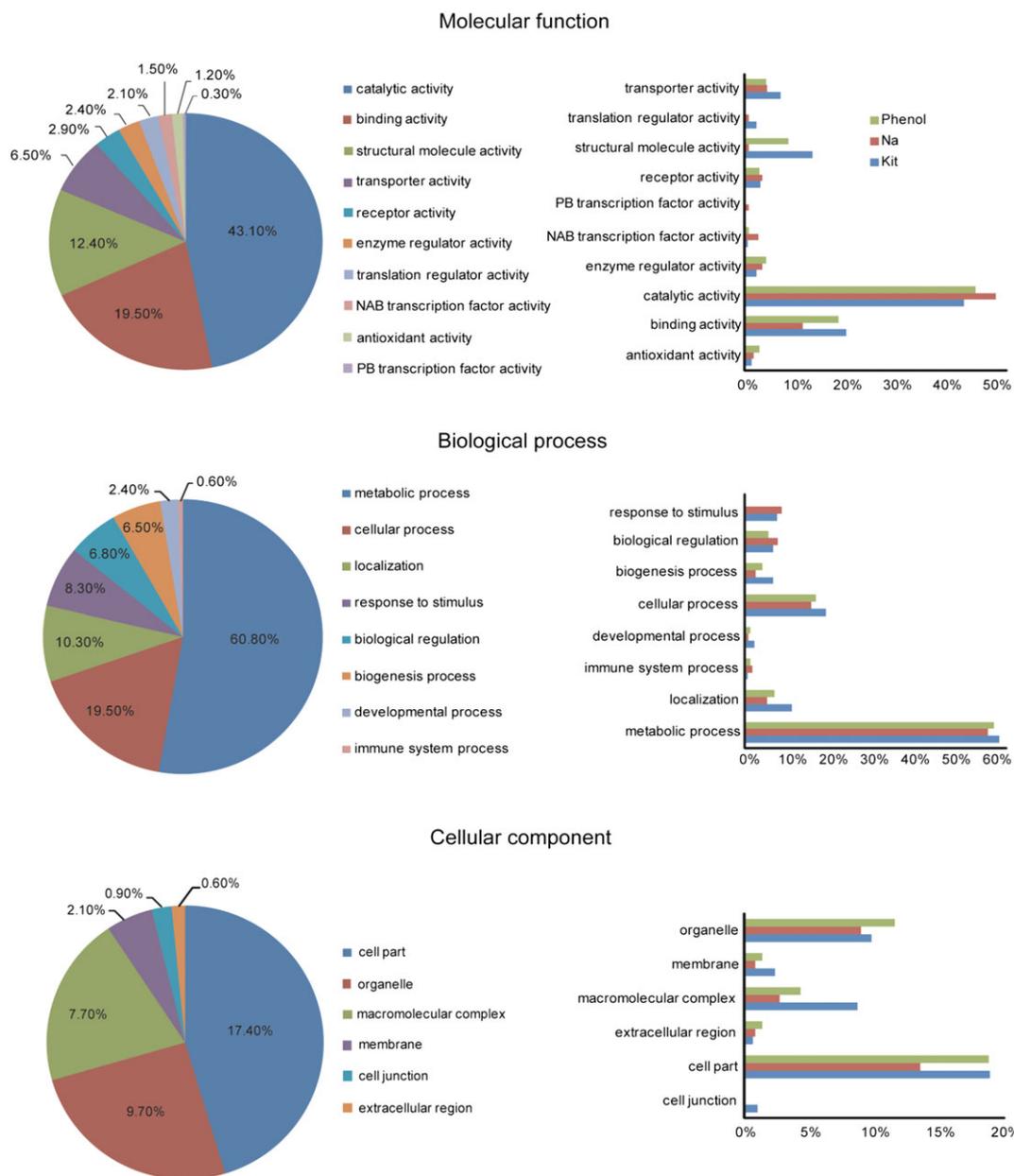


Figure 4. Functional classification of the identified proteins from cumin by orbitrap Velose mass spectrometer. Left side: Gene ontology pie chart of the total identified proteins are grouped into three categories: molecular function, biological process and cellular components. Right side, histogram presentation of gene ontology proteins derived from each extraction method. The x-axis indicates the percentage of gene against total number of genes and y-axis indicates the category name. [Colour figure can be viewed at wileyonlinelibrary.com]

organelle (10.50%) and macromolecular complexes (7.70%), while some were from membrane (1.8%), and a few from cell junction (0.4%) (Fig. 5).

Protein class analysis using PANTHER categorised proteins into 23 different groups in which oxido-reductase (15%), nucleic acid binding (13%), hydrolases (13%) and transferases (10%) comprised the largest group (Supporting Information, Fig. S2). Whereas, other proteins including proteases, immunity proteins, enzyme modulators and signalling molecules are less than 5% of total number of identified proteins. These proteins were mostly concerned with catalytic and binding activity.

Interestingly, along with a high number of housekeeping proteins, different forms of cruciferins have also been identified.

Cruciferins or 12S seed storage proteins belong to the 11S group of globulin. Although these proteins are highly abundant in the seed they are always identified in low numbers by MS due to the presence of multiple forms. Data shows that these cruciferins comprised only 0.1% of all identified protein. However, large amounts have been confirmed by high iBAQ intensities (Supporting Information, Table S4 and Fig. S2), which is proportional to the abundance of protein.

Despite the lack of information on cumin proteins we could successfully identify flavone synthase I (Gebhardt *et al.*, 2007). Data showed the presence of seven unique peptides out of a total of 12 identified peptides. Flavone synthase I catalysed the formation of flavones that exhibit diverse biological properties including anti-

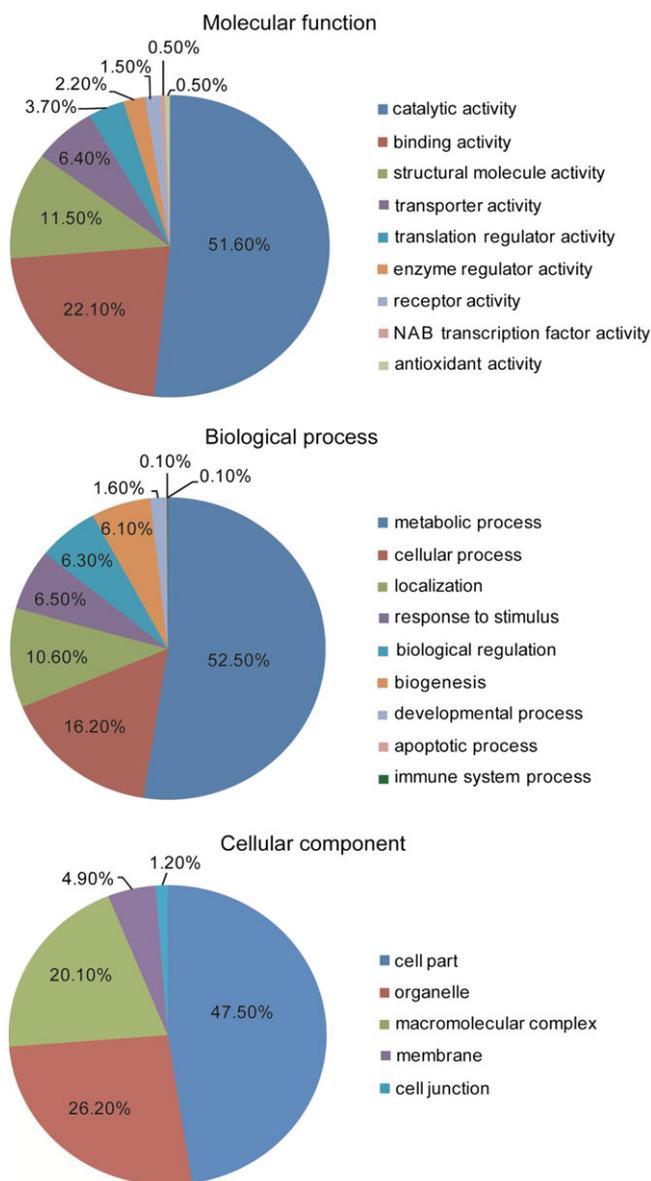


Figure 5. Functional classification of the identified proteins from cummin by Q-Exactive HF mass spectrometer orbitrap. Gene ontology pie chart of the identified proteins are grouped into three categories: molecular function, biological process and cellular components. [Colour figure can be viewed at wileyonlinelibrary.com]

oxidative, antitumor, anti-inflammatory, antibacterial, antiviral, and anti-atherosclerotic activities.

Conclusion

Cumin is popular not only as a spice but also in traditional medicine. A number of studies on cumin suggest the presence of biologically active constituents supporting its traditional use. Cumin contains a good amount of proteins, typically 10%. However, potential therapeutic properties of cumin protein remain poorly understood because no detail protein profiling of cumin has yet been reported. As protein solubility is a prerequisite for enhanced proteome analysis the first step in our proteome analysis strategy was evaluation of different extraction methods. Among

the three protein extraction methods, ReadyPrep kit provided high purity of protein extract for 1DE. Analysis by high mass accuracy, high speed Q-Exactive HF resulted in identification of a very high number of proteins. Further bioinformatics analysis provides functional distribution of identified proteins. Thus, in the present study, we have successfully applied 1DE and MS-based proteomics approach to generate a comprehensive data set of cummin proteome. Since the use of herbal drugs is becoming more popular in the form of food supplements, and in alternative medicine, we believe that this proteome analysis will provide valuable detail of protein constituents and help to further understand the nutritional and therapeutic properties of cummin.

Acknowledgements

The authors wish to express their gratefulness to Monika Rabbe and Uwe Plessman for the LC-MS/MS data collection, Sunit Mandar for MaxQuant data processing and Waqasuddin Khan for bioinformatics analysis.

The authors are very thankful to the Higher Education Commission of Pakistan for financial assistance wide Pak-US phase II project "Nanomedicine for Cancer Research".

Declaration of interest statement

The authors have no competing interests.

References

- Atanassov I, Urlaub H. 2013. Increased proteome coverage by combining PAGE and peptide isoelectric focusing: Comparative study of gel-based separation approaches. *Proteomics* **13**: 2947–2955.
- Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254.
- Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.
- Cox J, Neuhäuser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**: 1794–1805.
- Ekta Prakash DKG. 2014. Cytotoxic Activity of ethanolic extract of *Cuminum cyminum* Linn against seven human cancer cell line. *Universal J Agri Res* **2**: 27–30.
- El-Ghorab AH, Nauman M, Anjum FM, Hussain S, Nadeem M. 2010. A comparative study on chemical composition and antioxidant activity of ginger (*Zingiber officinale*) and cummin (*Cuminum cyminum*). *J Agri Food Chem* **58**: 8231–8237.
- Faurobert M, Pelpoir E, Chaib J. 2007. Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues. *Methods Mol Biol* **355**: 9–14.
- Gachkar LYD, Rezaei MB, Taghizadeh M, Astaneh SA, Rasooli I. 2007. Chemical and biological characteristics of *Cuminum cyminum* and *Rosemarinus officinalis* essential oils. *Food Chem* **102**: 898–904.
- Garber EA, Parker CH, Handy SM, Cho CY, Panda R, Samadpour M, Reynaud DH, Ziobro GC. 2016. Presence of undeclared food allergens in cummin: The need for multiplex methods. *J Agri Food Chem* **64**: 1202–1211.
- Gebhardt YH, Witte S, Steuber H, Matern U, Martens S. 2007. Evolution of flavone synthase I from parsley flavanone 3beta-hydroxylase by site-directed mutagenesis. *Plant Physiol* **144**: 1442–1454.
- Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* **97**: 9390–9395.
- Hajlaoui H, Mighri H, Noumi E, Snoussi M, Trabelsi N, Ksouri R, Bakhrouf A. 2010. Chemical composition and biological activities of Tunisian *Cuminum cyminum* L. essential oil: A high effectiveness against *Vibrio* spp. strains. *Food Chem Toxicol: An Int J Published British Ind Biol Res Assoc* **48**: 2186–2192.
- Jagtap AG, Patil PB. 2010. Antihyperglycemic activity and inhibition of advanced glycation end product formation by *Cuminum cyminum* in

- streptozotocin induced diabetic rats. *Food Chem Toxicol: An Int J Published British Ind Biol Res Assoc* **48**: 2030–2036.
- Masoumi SM, Kahrizi D, Rostami-Ahmadvandi H, Soorni J, Kiani S, Mostafaie A, Yari K. 2012. Genetic diversity study of some medicinal plant accessions belong to Apiaceae family based on seed storage proteins patterns. *Mol Biol Rep* **39**: 10361–10365.
- Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**: 1551–1566.
- Mnif S, Aifa S. 2015. Cumin (*Cuminum cyminum* L.) from traditional uses to potential biomedical applications. *Chem Biodivers* **12**: 733–742.
- Nomura Y, Arai Y, Shiomi N. 1999. Distribution and some properties of amylase in spices. *Food Sci Technol Res* **2**: 161–163.
- Ong SE, Pandey A. 2001. An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng* **18**: 195–205.
- Pan KT, Chen CC, Urlaub H, Khoo KH. 2017. Adapting data-independent acquisition for mass spectrometry-based protein site-specific N-glycosylation analysis. *Anal Chem* **89**: 4532–4539.
- Schmidt C, Urlaub H. 2009. iTRAQ-labeling of in-gel digested proteins for relative quantification. *Methods Mol Biol* **564**: 207–226.
- Schwanhaussner B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.
- Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dienes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H. 2014. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**: 223–226.
- Wang W, Tai F, Chen S. 2008. Optimizing protein extraction from plant tissues for enhanced proteomics analysis. *J Separation Sci* **31**: 2032–2039.
- Yan Zehn JS. 2011. Evaluation of sample extraction methods for proteomic analysis of coniferous seeds. *Acta Physiol Plant* **33**: 1623–1630.
- Zaman U, Abbasi A. 2009. Isolation, purification and characterization of a nonspecific lipid transfer protein from *Cuminum cyminum*. *Phytochemistry* **70**: 979–987.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.