The Major Proteins of the Seed of the Fruit of the Date Palm (*Phoenix dactylifera* L.): Characterisation and Emulsifying Properties

Ibrahim Akasha, Lydia Campbell, Julien Lonchamp & Stephen R. Euston*

Department of Food & Beverage Science, School of Life Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom

*Corresponding Author

Email: S.R.Euston@hw.ac.uk
Tel: +44 131 451 3640
Fax: +44 131 451 3009
Abstract

Proteins were extracted from the seeds of the fruit of the date palm. Proteomic analysis and SDS-PAGE electrophoresis of the extracted proteome suggested it is composed predominantly of the storage proteins glycinin and β-conglycinin, although over 300 proteins were detected, 91 of which were identified with confidence. In terms of protein type, the largest numbers of proteins were associated, not unexpectedly, with metabolism and energy functions, which reflected the requirements of the germinating and growing embryonic plant. The emulsifying properties of the extracted proteins were determined. Date seed protein exhibited a lower emulsifying activity than either whey protein concentrate or soy protein isolate at each of the pH values tested. However, the stability of the emulsions produced with all three proteins was very similar at the different pH values. This combination of large emulsion droplet size and high emulsion stability properties suggested that the date proteins may adsorb as large protein oligomers.

Keywords: Date seed proteins; proteomics; emulsifying properties
1. Introduction

The increasing cost of proteins from animal sources such as meat, egg and dairy products has encouraged the food industry to find alternative sources of proteins for use as functional ingredients in formulated foods. In addition it is becoming evident that protein sources such as fish meal or soy protein that are often used as animal feed are unsustainable or economically not viable. Plant proteins, such as soy, legume, canola and cereal proteins are appealing as sources of food protein because their production is more sustainable (Pimentel & Pimentel, 2003). However, plant proteins are often more difficult to extract, require large quantities of water during the extraction process and may lose functional properties during extraction (Schutyser & van der Goot, 2011). Loss of functional properties occurs due to loss of solubility brought about by denaturation of the protein under the extreme conditions (acid or alkaline and heating) required to extract the proteins from the polysaccharide-containing plant matrix (Schutyser & van der Goot, 2011).

The fruit of the date palm Phoenix dactylifera L. is one of the richest fruit-based sources of protein. Date palm is one of the major fruit crops produced in dry and semidry regions. It is an important commercial crop in different regions of the world (Al-Yahyai & Manickavasagan, 2012) and is considered the third most important palm species in the global agricultural industry after coconut and oil palms. The seeds of the date fruit, which are a waste product from date processing, also contain 5–7% protein by weight (Aldhaheri et al., 2004), but very little is known about the composition and the functional properties of these seed proteins. If it is possible to extract the proteins from the seeds it might be useful as a source of protein for human or animal nutrition. Robust methods for the extraction of proteins from date
seeds could facilitate the utilisation of date palm wastes such as seeds in the human and animal diet.

Functional properties of proteins define their behaviour in a food system during production and processing. Extraction and isolation of proteins from plant seeds is only the first step to integrating these proteins into food products. If they are to be of use as food ingredients they have to prove sufficiently functional to be used in place of current food proteins such as milk, egg and soy proteins. Studies of the functional properties of new protein sources can provide valuable information on the potential effectiveness of the proteins in food products. The important functional properties of proteins in food applications are solubility, swelling and water / fat holding capacity, emulsifying activity and emulsion stability, foaming ability and foam stability and gelling capacity.

There is a lack of information in the literature on the functional properties of proteins from date palm seed. This study aimed to investigate the extraction of protein from date seed, characterise these proteins using mass spectrometry and test their emulsifying properties.
2. Materials & Methods

All chemicals were purchased from Sigma Aldrich, Dorset, UK unless stated.

2.1 Preparation of Date Seed Protein Isolate

Dates (i.e. the fruit of the date palm Phoenix dactylifera L.) were purchased from a local supermarket in Edinburgh, United Kingdom. The dates were purchased at the Tamr stage (complete maturity) and their variety was Deglet Nour that had been grown in Tunisia. Seeds were removed from 40 kg of whole dates, washed in water to remove any remaining date flesh and then air-dried for a week. The seed was found to make up 10.3% (w/w) of the total mass of the date fruit on average. The seeds were then further dried overnight at 40ºC in a drying oven. The seeds were milled using a hammer mill to a particle size that could pass through a 1–2 mm sieve screen and then stored at –20ºC until further preparation was required. The powder obtained was identified as date palm seed powder (DPSP). The composition (w/w) of the DPSP has been reported in our previous paper as protein, 5.64%, moisture, 5.39%, fat 8.14%, fibre 18.50%, ash 0.95%, carbohydrate 61.38% (Akasha, Campbell & Euston, 2012).

Oil was extracted from DPSP using a Soxhlet apparatus. Fifteen gram samples of dried DPSP were weighed into an extraction thimble (Fisher Scientific, UK) and sealed with cotton wool. The thimble was inserted in a Soxhlet extraction flask and extracted with boiling hexane (boiling point 68 ºC) for 10 hours or until the solvent at the sample chamber was colourless, indicating it was free from oil and that all the oil had been extracted. The defatted DPSP was removed from the extraction thimble and left to dry overnight to allow the hexane to evaporate. This defatted date seed powder (DDSP) was kept at –20ºC until processed further. The residual fat content and protein content of the defatted powder were reported previously (Akasha,
as 1.01% (w/w) and 6.13% (w/w) respectively. This protein content is equivalent to a 100% yield of protein. The effect of the hexane extraction step on the functionality of the proteins was not determined. However, it is well known that the methods used to extract the proteins from the powder will also affect the functionality so the additional effect of hexane extraction is likely to be negligible.

2.1.1 Protein Isolation

Protein was extracted from the DDSP using a phenol/trichloracetic acid (Ph/TCA) extraction procedure based on the methods (with some modifications) proposed by Gomez–Vidal et al., (2008) for olive and Phoenix dactylifera L. leaves respectively. Ten grams of defatted DDSP was mixed with 30mL of ice–cold acetone, vortex mixed and then centrifuged at 10,000rpm for 10 min at 4°C (Beckman Avanti J26-XP centrifuge). The supernatant was decanted and discarded and the residual pellet washed twice with ice–cold acetone and allowed to dry at room temperature. After the pellet had dried it was ground to a fine powder using a pestle and mortar, rinsed with 15% (w/v) TCA in acetone, vortex mixed and then centrifuged at 10,000 rpm for 10min at 4°C. The rinsing with TCA/acetone and centrifugation was repeated three times. The pellet was then rinsed with cold 15% (w/v) TCA in water and centrifuged. The rinsing with cold TCA and centrifugation was repeated three times. The pellet was then rinsed with cold 80% (v/v) acetone followed by centrifugation, and this was also repeated three times. The pellet was then air dried.

2.1.2 Protein Purification
To purify the protein two grams of the dry pellet was suspended in a mixture of 10 mL of Ph/Tris–buffer, pH 8.0 and 10 mL of dense SDS buffer (2%[w/v] SDS, 5%[w/v] sucrose, 0.1M Tris–HCL, pH 8.0, 5% [v/v] β–mercaptoethanol). The mixture was vortex mixed and the pellet was obtained by centrifugation at 10,000 rpm for 10 min at 4°C using a Beckman Avanti J26-XP centrifuge fitted with a JA25.50 rotor (Beckman-Coulter, High Wycombe, UK). The pellet was resuspended in Ph/Tris–buffer and dense SDS solution, and centrifuged again under the same conditions.

The pellets from both centrifugations were mixed and precipitated with five volumes of cold 0.1M ammonium acetate in methanol, refrigerated at 4°C overnight and then centrifuged at 10,000 rpm for 10 min at 4°C. The pellet from this centrifugation was then washed three times with cold methanol plus 0.1M ammonium acetate and centrifuged as above followed by the same process with cold 80% (v/v) acetone. Half a gram of the dried pellet was then mixed with 5 mL of cold aqueous 24% (w/v) TCA, vortex mixed and left to precipitate on ice for 30 min, followed by centrifugation at 13,000 rpm for 15 min at 4°C. The pellet was washed with 2 mL of ice cold acetone, incubated for 15 min on ice and then centrifuged at 13,000 rpm for 15 min at 4°C. The final pellet or date seed protein concentrate (DSPC) was air-dried in an oven at 30°C overnight (16 hours) and stored at −20°C until required for further analysis.

2.1.3 Protein Content of DSPC

The crude protein content of the extracted DSPC and DDSP was determined by measurement of the nitrogen content using the Kjeldahl method (Lynch, Barbano & Fleming, 1998).
The percent yield of protein from the date palm seed was determined by calculating the protein recovered in the DSPC and comparing this to the maximum possible protein recovery from the DDSP.

2.2 SDS-PAGE Analysis of DSPC

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) was performed on date palm seed on a 12% polyacrylamide gel (BioRad, Hemel Hempstead, UK). A sample of DSPC and soy protein isolate (SPI) were run on the gel. A protein molecular weight ladder (BioRad, Hemel Hempstead, UK) was also run on the gel to allow molecular weight determination. Gels were stained overnight with colloidal Coomasie blue and destained (10% [v/v] ethanol and 2% [v/v] orthophosphoric acid) until the background became clear and protein bands were visible. Gels were scanned using a BIO-RAD Molecular imager® (ChemiDocTM XRS+) and analysed using GelAnalyzer 2010a software to estimate the molecular weight of protein bands.

2.3 Preparation of Protein for LC-MSMS

Protein preparation was carried out using a method proposed by Le Bihan et al., (2011). Ten mg of DSPC was resuspended in 50 µL of distilled water (dH2O), followed by denaturation with 250 µL of 8M urea and dilution with 950 µL dH2O prior to trichloroacetic acid (TCA) precipitation with 310 µL of 100% TCA, 1250 µL methanol and 625 µL chloroform. Samples were vortex-mixed and incubated (4°C, 10 min) before centrifugation (4,500g, 4 °C, 10 min). The top phase was removed before adding 1 mL methanol. The sample was vortex mixed before centrifugation
(4,500xg, 4 °C, 10 min), the supernatants were removed and the solid sample washed twice with 1 mL acetone, centrifuged at 10,000g, at 4 °C for 5 min and dried under vacuum. Then, the sample was resuspended in 100 µL dH2O.

Protein digestion was carried out using the method proposed by Le Bihan et al., (2011) on 20µL of protein extract. Briefly, samples were denatured in 8M urea, reduced by incubating with dithiothreithol (DDT) prior to cysteine alkylation with iodoacetamide and overnight digestion with 60 µg trypsin at room temperature. Four µg of peptide samples were acidified with 1% formic acid before centrifugation and cleaning using Stage tips (Thermo Scientific, Hemel Hempstead, UK). Finally, the peptide samples were vacuum–dried and stored at –20 °C until further analysis.

Two µg peptide samples were analysed in a randomised sequence by capillary HPLC–MSMS, using 140-minute gradients as described by Martin et al. (2012), on an on-line system consisting of a micro-pump (1200 binary HPLC system, Agilent, UK) coupled to a hybrid LTQ--Orbitrap XL instrument (Thermo-Fisher, UK). HPLC quality acetonitrile (Fisher, UK) and water were used. Suprapure 98–100% formic acid and 99% purity sequencing grade trifluoroacetic acid were purchased from Merck (Darmstadt, Germany).

2.3.1 Identification and Quantification of the Peptides

Multicharged (2+, 3+ and 4+) ion intensities were extracted from the LC-MS files and the Mascot Version 2.4 software (Matrix Science Ltd, UK) was used to compare the MSMS data against the NCBI protein database (13/03/2013; 11,961,441 sequences). Search parameters used were a maximum missed-cut value of 2, variable oxidation (M), N–terminal protein acetylation and fixed carbamidomethylation (C), precursor mass tolerance 7 ppm and MSMS tolerance
0.4 Da. A significance threshold (p) of <0.05 (MudPIT scoring) was set and a
minimum peptide cut off score of 20. Proteins identified and quantified with 2 or more
peptide sequences were retained.

2.4 Emulsifying Properties
Emulsifying activity index (EAI) and emulsion stability index (ESI) were determined
by a turbidimetric method according to Ogunwolu et al. (2009) with some
modifications. Four hundred and fifty milligrams of protein sample was dispersed in
45 mL of Mill–Q water. The protein solution was then mixed with 15 mL of sunflower
oil purchased from a local supermarket (Tesco Ltd, UK) and the pH was adjusted to
2, 4, 6, 8, 10 or 12 using 0.1M HCl or 0.1M NaoH. The mixture was homogenised
using an Ultra–turrax high speed homogenizer (IKA–Werke GmbH, Germany) for 1
min to make a protein–stabilised oil–in–water emulsion. Fifty µL of the emulsion was
removed from the bottom of the container using a pipette and suspended in 5 mL of
0.1% (w/v) SDS solution. This was carried out immediately at 0 min and 10 min after
the homogenisation. Absorbance of the diluted emulsions was measured at 500nm
using a UV/Vis spectrophotometer (Model Genesys 6, Thermo Electron Corporation,
USA). The ability of the protein to form an emulsion (emulsifying activity index, EAI)
and the stability of the formed emulsion (emulsion stability index, ESI) were
calculated using the following formulae:

\[
\text{EAI (m}^2/\text{g)} = \frac{2 \times T \times A_0 \times \text{dilution factor}}{C \times \phi \times 1000}
\]

\[
\text{ESI (min)} = \frac{A_0}{A_0 - A_{10}} \times \Delta t
\]
Where, $T = 2.303$, $A_0 =$ absorbance immediately after the homogenisation, dilution factor $= 100$, $C =$ the weight of protein per unit volume (g/mL), $\phi =$ the oil volumetric fraction (0.25), $A_{10} =$ absorbance after 10 min of the homogenisation, $\Delta t = 10$ min.

The emulsifying ability and emulsion stability was repeated in triplicate and the error bars quoted as the standard deviation of the mean.

3 Results & Discussion

The DSPC powder obtained showed a crude protein content of 68% (w/w) and 44% of the protein was recovered from the defatted date seed powder. This DSPC was used for subsequent proteomic analysis and functional testing.

3.1 Identification of the Date Palm Seed Protein Isolates by LC-MSMS

Over three hundred proteins were detected in the DSPC sample by LC-MSMS. Not all identifications were considered significant (see below). Protein identification was achieved after the MSMS data were compared to known sequences on the NCBI database using the Mascot Version 2.4 software (Matrix Science Ltd, UK). This search resulted in 318 hits, each of which corresponding to a unique protein. The protein list was screened to remove any contaminants (e.g. proteins that the database only identified as being found in humans or animals). Since the preparation method for the LC-MSMS requires digestion of the sample with trypsin, this protein, corresponding to the hit number 1 (i.e. the most abundant protein) is ignored A second protein, keratin (hit number 59), an animal protein found in hair, nails and skin, was also removed as this was considered to be a contaminant. To determine how accurate the identification of the remaining proteins was we used two criteria, the MOWSE score and the condition that the identification be based on at least two peptides being matched to the predicted peptide map of the protein. MOWSE
(Molecular Weight Search) is a method that aids in identifying proteins based on molecular weight of the peptides formed from proteolytic digestion of the protein sample by allowing the probability of correct identification of the protein to be calculated. The method was first developed by Pappin, Hojrup & Bleasby (1993). This method calculates the probability that the peptide has been misidentified during database searching, i.e. the identification is a random event. A low probability (P) of misidentification is required for correct identification. Since it is more common to express a more accurate identification as a higher number, the probability of misidentification is converted to a MOWSE score using the formula,

$$MOWSE\ Score = -10 \log_{10}(P) \quad (1)$$

For example, using equation (1), protein identification with a probability of $10^{-10}$ that it is a misidentification will have a MOWSE score of 100. The probability is calculated based on the number of peptide matches identified for a particular protein match compared to the sequence database using an algorithm detailed by Pappin, Hojrup & Bleasby (1993). To determine whether a particular MOWSE score is significant, a cut-off value is defined based on the assumption that a random event is acceptable if it occurs less than 5% of the time. To calculate the cut-off MOWSE score we need to calculate the probability of a random event across the whole of the protein database that is searched for matches. At the time the LC-MSMS results were submitted the NCBI protein reference database contained 11,961,441 sequences. A 5% probability of a random identification is equivalent to 1 in 20 mismatches, so the MOWSE cut-off score will be:

$$MOWSE\ cut-off = -10 \log_{10} \left( \frac{1}{20 \times 11961441} \right) = 83.7 \quad (2)$$
Therefore any protein match with a MOWSE score of 84 or greater will have less than a 5% chance of being an incorrect identification. The first 111 hits were considered to have been successfully identified since they all displayed a MOWSE score of 85 or greater and therefore can be considered to be found in date palm (Phoenix dactylifera L.) seed. However, on closer inspection not all of these have been identified as a particular protein, with some being labelled unknown proteins, and some hypothetical (identified from gene sequences) but which are nonetheless in the NCBI database. Other proteins failed the second criterion that more than one peptide is used in the identification. Once these proteins had been removed along with contaminants, 90 unique proteins were identified. These 90 most abundant proteins were classified into twelve different groups according to their functions using the categories described by Bevan et al (1998). The different functional group classifications and percentages found in the DSPC are show in Table 1.

Three of the groups (groups 3, 8 and 9) have no representative proteins identified amongst the 90 proteins. Several of the proteins identified have previously been reported before and have known functions. A table listing all 90 identified proteins is available as supplementary material.

The twenty most abundant proteins are listed in Table 2. Data listed in Table 2 include the hit number (HN), protein description, molecular weight search score (MOWSE score), protein molecular weight (MW) and number of peptide matches compared to total number of peptides produced. The hit number is a rough indicator of protein abundance in the sample, with a higher hit number indicating a more
abundant protein. A discussion of the function of the twenty most abundant proteins follows according to the functional category they belong to.

**Functional category 1:** Lipoxygenase was identified in this category with a hit number of 4. Lipoxygenase is an iron-containing enzyme that catalyses the formation of hydroperoxides in fatty acids that contain a pentadiene segment (Andreou & Feussner, 2009). Functional properties of lipoxygenase in foods have not been reported. However, it is known that lipoxygenase catalysed formation of peroxide free radicals can promote the crosslinking of soy proteins, reducing solubility and adversely affecting functional properties such as gelling ability (Kong, Li, Wang, Hua & Huang, 2008). A second protein from this category, β-amylase, was identified with a hit number of 10. This enzyme is found in plant seeds that have starch as the primary storage polysaccharide. In plant seeds it functions to break down starch into maltose when carbohydrate is required for glycolysis during plant growth (Smith, Zeeman & Smith, 2005). There are no reports of its functional properties in food, other than as an enzyme, although its ability to form foams is evidenced by the use of foam fractionation in its separation (Nakabayashi, Takakusagi, Iwabata & Sakaguchi, 2011).

**Functional category 2:** Proteins in this category are involved with energy metabolism in the cell, and the high abundance of these proteins reflects the high energy requirements required in a germinating and growing embryo plant.

Ribulose-1,5-bisphosphate carboxylase (RuBisCo) (hit no. 8) is one of the most abundant proteins on Earth being found in all green plants. The biological function of RuBisCO is to catalyze two reactions: the carboxylation of D-ribulose 1,5-bisphosphate, the primary event in carbon dioxide fixation and the oxidative fragmentation of the pentose substrate in the photorespiration process. The potential
of Rubisco as a food protein has been hypothesized for many years (Douillard & de Mathan, 1994). Recently the focus has been on the extraction of rubisco from the leaves of green plants, and this has revealed that, depending on the extraction method, rubisco powders with good functional properties can be made (Kamm, Kamm, Scherze, Muschiolik & Binbrich, 2006).

**Functional category 4:** The most abundant protein in category 4, EM1 was not one of the twenty most abundant proteins with a hit no. of 51, However, it was one of the few proteins that was positively identified in the NCBI database as being from *Phoenix dactylifera* L. EM1 is one of the stress induced proteins that are expressed in times of drought to protect cells from dehydration stress at the molecular level (Sham & Aly, 2012).

**Functional category 6:** In this category several proteins were identified in the 20 most abundant, glycinin (hit no. 2); alpha subunit of beta conglycinin (hit no. 3); chloroplast protein precursor LI818R (hit no. 11); heat shock cognate 70 kDa protein (HSP70) (hit no. 16). Glycinin and beta conglycinin are the two most abundant proteins identified in our date seed sample. To confirm this SDS-PAGE electrophoresis was carried out to assess the molecular weight profile of the major proteins. SDS-PAGE of DSPC was undertaken under reducing and non-reducing conditions. Soy protein isolate was also run on the gels since this is known to be comprised mainly of glycinin and conglycinin. Pictures of the SDS-PAGE gels are shown in Figure 1, and the results are summarised in Table 3.

The most abundant protein band occurred at 60kDa, with minor bands identified at higher and lower molecular weights for the date seed protein isolate. Using non-reducing conditions (data not shown) did not alter the protein band profile significantly, suggesting that disulphide bonds were absent from these proteins. It
was previously reported that albumins of oil palm seeds did not display disulphide
bonds (Morcillo et al., 1997). Khoshroo et al. (2011) reported similar results based on
an analysis of seed protein from twelve varieties of date palm (Bazmani sefid (Bw.Ji),
Mahminai, Gordial, Kharok, Almehtari, Mordar sang, Kaluteh, Halilehi, Bazmani sefid
(Bw.Ba), Mazafati, Khorbak syah, Khosh kang) grown in different Iranian regions.
The researchers found one heavily stained band at around 65kDa and minor bands
ranging from 12 to 369 kDa. Bouaziz et al. (2008) found three similar prominent
protein bands in date seeds of Allig and Deglet Nour varieties at 32, 60 and 70KDa.
The differences in protein profile between our results and the previous work (Bouaziz
et al., 2008) could be explained by a number of factors. The extraction process used
in the other studies differs from ours and this may lead to differential extraction of
proteins. Variation between the seed storage proteins is expected within different
varieties of the same species. In particular, extensive genetic polymorphism of seed
proteins is observed both within the same genotype and among genotypes of the
same species. This genetic polymorphism may occur through the presence of
multigene families within the same species, or through post-translational
glycosylation of proteins or proteolytic action on the proteins (Miernyk and Hajduch,
2011). Glycosylation, in particular, will lead to several proteins with the same amino
acid sequence but differing molecular weight due to the presence of one or more
sugar chains of variable length and position. Finally, the protein composition of the
seed varies during the embryo development process, with the major storage protein
not appearing until 3 months after fertilization. Thus, the level of maturity of the date
fruit will also influence the protein profile found in the seed. This may partly explain
the differences in molecular weight profile for the seeds proteins found in our study
and those of Bouaziz et al. (2008), Ehsanpour et al. (2010). Purification and
characterization of storage proteins in oil palm embryo (the same family as the date palm) has been studied by Morcillo et al, (1997). They identified the major storage proteins as being 2S and 7S globulins. The 2S proteins were made up of two polypeptides (one acidic and one basic) of 22 kDa and 19 kDa molecular weight respectively. The 7S proteins were the major fraction identified using SDS-PAGE. These were shown to be a heterogeneous group of polypeptides of molecular weight between 45 and 65 kDa with no disulphide bonds. They were also found in the form of oligomers with molecular weights of 156 and 201 kDa.

For comparison purpose a soy protein isolate sample was also run on an SDS PAGE gel. Six intense, detectable bands were observed (lane C, Figure 1), located at approximately 535, 64, 50, 36, 22 and 16 kDa respectively. These bands might be identified with basic polypeptides of glycinin which have an accepted molecular weight range from 16-22KDa, acidic polypeptides of glycinin with molecular weight range 34-36KDa, β-subunit (40-50KDa) and α-subunit (64KDa) (Roesch & Corredig, 2005). The high molecular weight band at 535 KDa could correspond to oligomers of glycinin.

Glycinin and conglycinin are known to be major storage proteins in most seeds, and in particular in soy beans (Utsumi, Matsumura, & Mori, 1997). The relationship between the molecular and functional properties of glycinin and beta conglycinin subunit has also been investigated in several studies (Maruyama et al., 2004; Utsumi, Katsube, Ishige & Takaiwa, 1997). It has been found that beta conglycinin has very good emulsifying properties and is a better emulsifier than glycinin (Molina et al., 2001). This is due to beta conglycinin having a larger number of hydrophobic groups with higher molecular flexibility compared to other protein fractions (Bernard
et al., 2001). The functional properties of these two proteins will be discussed further below.

**Functional category 11:** Dakhlaoui-Dkhil et al. (2013) report that 16.6% of the identified proteins of date palm (*Phoenix dactylifera* L.) leaf are defence-related proteins which include defence regulated proteins and resistance proteins, those involving detoxification, stress responses, cell rescue and cell death (Bevan et al., 1998). This compares to 7% of date seed proteins in this category found in this study (Table 1). The protein from this category which is found in the 20 most abundant was the seed biotin-containing protein (hit no. 7).

**Functional category 12:** There are several proteins in the 20 most abundant that are unidentified or tentatively identified. These include an unnamed protein product at hit no. 6; an unnamed protein product (hit no. 12) (possibly 7S globulin basic subunit); an unknown protein (hit no. 18) (possibly formate dehydrogenase); putative histone H2B (hit no. 9).

The NCBI database of proteins which was searched using Mascot during the analysis of the proteomics results is the largest store of experimentally identified biological macromolecular structures available. However, in this database there are a large number of proteins that have uncharacterized functions. Unnamed or hypothetical proteins are often those that have been identified based on genome sequencing of an organism, but the protein for which the gene codes has not been identified, named and characterised in the plant or animal. Dakhlaoui–Dkhil et al (2013) reported that 29.4% of protein detected in date palm leaf was hypothetical protein, not dissimilar from the 22% detected here (Table 1).
3.2 Emulsifying Activity and Emulsion Stability of Date Seed Proteins

Compared to SPI and WPC

For DSPC to be exploited as a food ingredient it must show comparable functional properties to other food proteins. The emulsifying properties (emulsifying ability (as emulsifying activity index (EAI) and emulsion stability as emulsion stability index (ESI)) were compared to soy protein isolate (SPI) and bovine whey protein concentrate (WPC) in Figures 2 and 3. SPI is a common plant protein emulsifier, and WPC is a highly functional animal protein emulsifier in formulated foods (Euston & Hirst, 2000). The emulsifying properties were tested over a range of pH. At all pH values the EAI of WPC was significantly greater than that of SPI which was in turn significantly greater than that of DSPC (Figure 2). All three samples showed a minimum in EAI at pH 4-5. This minimum occurred at the isoelectric point for both WPC at pH 4.8 (Demetriades, Coupland & McClements, 1997) and soy proteins between pH 4.7-5.0 (Golubovic, van Hateren, Ottens, Witkamp, van der Wielen, 2005). The same trends as seen for EAI between the three samples were not observed with the emulsion stability. In Figure 3 the ESI proved very similar for all three protein samples across the whole pH range. Furthermore, WPC emulsion ESI was slightly less than for DSPC at most pH values.

The size of emulsion droplets is a major factor in the stability of the emulsion, with larger droplets proving less stable than smaller droplets. Therefore, a correlation might be expected between EAI and ESI since the EAI is an indirect measurement of the droplet size. When the ESI is plotted against EAI a linear relationship between ESI and EAI for all three protein samples is observed as expected, i.e. a larger EAI (smaller particles size) resulted in more stable droplets (Supplementary Figure 1). Differences in the relationship between EAI and ESI are observed between the three
protein samples. EAI values were similar for both DSPC and SPI, however the slope of the EAI vs ESI graph was greater for the DSPC emulsions than for the SPI emulsions, suggesting that for a given droplet size the SPI emulsions were less stable. For WPC emulsions the EAI was high compared to DSPC and SPI emulsions, but the ESI was lower for a given EAI, although the correlation was still linear with a slope very similar to that for the SPI.

Soy bean protein emulsifying functionality has been widely studied (Utsumi, Katsumura, & Mori, 1997). Soy proteins are predominantly glycinin and β-conglycinin (70% of the total protein) and these two proteins determine the emulsifying properties. The DSPC was shown above to contain high levels of glycinin and β–conglycinin so we would expect these proteins to play a large part in the emulsifying behaviour of DSPC.

The quaternary structure of both glycinin and β-conglycinin is complex. In the plant seed, glycinin is found as a hexamer (molecular weight in the range 300–380 kDa), and is made up of combinations of 5 distinct subunits (Staswick, Hermodson, Nielsen, 1984). Glycinin hexamers can form trimers (7S) or monomers (3S) by dissociation at different pH and ionic strength combinations (Peng, Quass, Dayton & Allen 1984). β-conglycinin also forms oligomers comprised of three polypeptide chains (α, α′ and β) with overall molecular weight in the range 150–200 kDa (Thanh & Shibasaki, 1979). The subunit composition of β–conglycinin is also variable. Soy proteins have been found to form adsorbed layers 30–40 nm thin at the surface of oil droplets (Keerati–u–rai & Corredig, 2010). Whey proteins such as β-lactoglobulin, on the other hand, form adsorbed layers that are only 4–6 nm thick (Atkinson, Dickinson, Horne & Richardson, 1995). The conclusion that can be drawn is that soy
proteins adsorb as aggregates (oligomers) rather than individual proteins unlike β-lactoglobulin. Maruyama et al. (2004) found that the subunit composition of the hexameric glycinin affects the emulsifying properties. Since the glycinin subunit composition is variable (Staswick, Hermodson, Nielsen, 1984) the emulsifying ability of soy proteins may vary. The subunit composition of β–conglycinin also affects emulsifying ability (Utsumi, Matsumura & Mori, 1997). The α subunit has been sown to be the best emulsifier followed by α' and then β (Utsumi, Matsumura & Mori, 1997). In addition, β–conglycinin is a better emulsifier than glycinin, due to its ability to adsorb more rapidly at the emulsion droplet surface and to spread more extensively at the interface (Utsumi, Matsumura & Mori, 1997; Bernard, Grandison & Lewis, 2001; Molina, Papadapoulou & Ledward, 2001). Clearly, the relative proportion of glycinin and β-conglycinin and their subunit composition affected the emulsifying properties of the DSPC and SPI powders, and this could explain the differences in emulsifying properties between the DSPC and SPI. Distinct differences in the protein molecular weight profile between the DSPC and SPI was observed in the SDS-PAGE results with a greater proportion of high molecular weight protein fractions seen in the DSPC (Table 3 and Figure 1).

We can speculate as to why DSPC emulsions are more stable than SPI emulsions of the same EAI (Supplementary Figure 1) by considering what is already known about the emulsifying properties of aggregated proteins. It has been observed previously (Euston & Hirst, 2000) that aggregated proteins are often poorer emulsifiers than non-aggregated proteins. However, the aggregated proteins emulsions displayed a greater stability under certain conditions. The explanation given for this was that the aggregates display a greater conformational stability than the native proteins, and were unable to unfold and spread rapidly to stabilise the oil droplet surface of small
droplets, which leads to larger droplets (lower EAI). On the other hand, because the proteins are in the form of large aggregates the adsorbed protein layer around the emulsions droplets is very dense and occupies a large volume, and leads to an increased emulsion droplet stability. Euston & Hirst (2000) proposed that two mechanisms were responsible for this increased emulsion stability. The density of the emulsion droplet is increased by the presence of the dense aggregated protein layer and this will reduce their creaming velocity and increase stability to creaming (Euston & Hirst, 2000). In addition, the extensive aggregated protein adsorbed layer is likely to increase the steric stabilising effect of the protein layer, thus reducing the likelihood of coalescence (Euston & Hirst, 2000). We have seen the presence of large oligomers of proteins in our DSPC, and in SPI (Table 3 & Figure 1) and this may explain the higher stability of DSPC emulsions over SPI and WPC emulsions. WPC proteins do form oligomers, but these are only loosely associated (Iametti, Scaglioni, Mazzini, Vecchio & Bonomi, 1998) and easily break up under emulsification conditions so that only protein monomers adsorb and a thin monomeric layer of protein is adsorbed to the emulsion droplet surface. This will have a lower steric stabilising ability, and lower effect on droplet density than the aggregates found in DSPC (and SPI).

The DSPC sample contains 32% non-protein which is almost certainly complex carbohydrate. We have carried out unpublished studies using various enzymes to aid the extraction of the protein. These suggest that there are high proportions of mannans, beta-glucans, xylans and cellulose present in the seeds, and that the seed proteins are more closely associated with the glucans and cellulose. Sekhar and DeMason (1988) have found that 75% of the protein in date palm seeds is found in the cotyledon parenchyma cells (part of the embryo), whilst only 17% is found in the
endosperm, where the mannans are found. We would therefore expect the date seed proteins to be associated with glucans, xylans and cellulose rather than mannans. The presence of these polysaccharides in the protein sample will influence the functional properties of the proteins. Recently, Bouaziz et al. (2013) have studied the functional properties of fibro-protein complexes from date seed and have demonstrated that they have potential as emulsifiers in food applications.

4 Conclusions

In this study we have extracted protein from the seeds of the date fruit and characterised them using proteomic analysis. LC–MSMS revealed a large number of proteins in the date seed protein sample. Of the 90 proteins identified with high confidence (MOWSE score above 84) the majority of these proteins (70% by number) have metabolic functions in the seed and seedling, whilst of the remainder 15% (by number) are storage proteins such as 11S and 7S globulin (glycinin and β-conglycinin) (Table 1). The emulsifying properties of DSPC were determined and it was found to have a comparable to SPI.

These results suggest that there is potential for DSPC as a functional ingredient in food systems. There are several factors to be considered when assessing a potential new protein source. The major factors are whether the protein can be isolated easily and cost-effectively in high enough quantities, and whether it displays the necessary functionality to replace other plant or animal proteins. Before date seed protein can be considered for use in foods these two factors would need to be addressed. The extraction process used in this study is not suitable for large scale food-grade extraction, and thus a procedure would need to be devised to extract the proteins in a food-grade manner. Secondly, all functional properties, not just emulsification, but
also foaming and gelation, will need to be characterised over a wider range of conditions that are relevant to food systems.

Acknowledgements

IA acknowledges the receipt of a PhD Scholarship from the Libyan government. Mass spectrometry was carried out at the Kinetic Parameter Facility, University of Edinburgh, UK.

5 References


Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R.,
Dirske, W., Van Staveren, M., Stiekema, W. et al. (1998). Analysis of 1.9 Mb of

Bouaziz, M., Besbes, S., Blecker, C., Wathelet, B., Deroanne, C., & Hamadi, A.
(2008). Protein and amino acid profiles of Tunisian Deglet Nour and Allig date palm

Dakhlaoui-Dkhil, S.S., Coquet, L., Cosette, P., Elkahoui, S., Song, P.C.T., Vaudry, D.,
dactylifera* L.) leaf proteome: identification of a gender biomarker to screen male

whey protein stabilized emulsions as related to pH and NaCl. *Journal of Food

In B.J.F. Hudson (Ed), New and Developing Sources of Proteins (pp 307–342). New
of seed storage protein patterns of four Iranian pistachios using SDS–PAGE. *Natural

emulsifying properties of protein products containing aggregated and non-


glycinins with enhanced food qualities and development of crops producing such 
glycinins. *Advances in Experimental and Medical Biology, 415*, 1–15.

and identification of soybean leaf proteins by two-dimensional gel electrophoresis 
Table Legends

Table 1 - The percentage of the 90 identified date seed proteins related to the functional categories identified by Bevan et al. (1998).

Table 2 - Twenty most abundant date palm seed proteins identified by Liquid-chromatography coupled mass spectrometry (LC-MSMS). HN = Hit number; MOWSE score as defined by equation (1); MW = protein molecular weight in Daltons (Da); Peptides matched = number of peptides matched/total number of peptides found.

Table 3 - Summary of the molecular weight of the protein bands identified in reduced SDS–PAGE gels from Figure 2. DSPC = Date seed protein concentrate; SPI = soy protein isolate.

Figure Legends

Figure 1 – SDS-PAGE results for date seed protein concentrate (DSPC) and soy protein isolate (SPI). Lane B is for DSPC and lanes C for SPI. Lane A contains a molecular weight marker with the molecular weights of the reference proteins marked.

Figure 2 – Emulsifying activity index (EAI) as a function of pH for date seed protein concentrate (DSPC), soy protein isolate (SPI) and whey protein concentrate (WPC). Error bars are ± one standard deviation of the mean.
Figure 3 – Emulsion stability index (ESI) as a function of pH for date seed protein concentrate (DSPC), soy protein isolate (SPI) and whey protein concentrate (WPC).

Error bars are ± one standard deviation of the mean.
<table>
<thead>
<tr>
<th>Protein Functional Category</th>
<th>Proteins in this Category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism/sugars and polysaccharides/amino acids/Nucleotides/Lipid</td>
<td>15</td>
</tr>
<tr>
<td>Energy/ ATP synthase/ Glycolysis/ Electrontransport/ Gluconeogenesis/ Photosynthesis/ Pentose phosphate</td>
<td>33</td>
</tr>
<tr>
<td>Cell growth/ division</td>
<td>0</td>
</tr>
<tr>
<td>Transcription/ mRNA</td>
<td>1</td>
</tr>
<tr>
<td>Protein synthesis/ Translation factors</td>
<td>8</td>
</tr>
<tr>
<td>Protein destination and storage/ Storage protein</td>
<td>10</td>
</tr>
<tr>
<td>Transporters/ Transport ATPases</td>
<td>3</td>
</tr>
<tr>
<td>Intracellular traffic</td>
<td>0</td>
</tr>
<tr>
<td>Cell structure</td>
<td>0</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>1</td>
</tr>
<tr>
<td>Stress responses/Disease/defence/pathogenesis-related protein</td>
<td>7</td>
</tr>
<tr>
<td>Unclear classification</td>
<td>22</td>
</tr>
<tr>
<td>HN</td>
<td>Proteins description</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>2</td>
<td>Glycinin</td>
</tr>
<tr>
<td>3</td>
<td>alpha subunit of beta conglycinin</td>
</tr>
<tr>
<td>4</td>
<td>Lipooxygenase</td>
</tr>
<tr>
<td>5</td>
<td>Sucrose-binding protein</td>
</tr>
<tr>
<td>6</td>
<td>unnamed protein product</td>
</tr>
<tr>
<td>7</td>
<td>Seed biotin-containing protein</td>
</tr>
<tr>
<td>8</td>
<td>ribulose-1,5-bisphosphate carboxy.</td>
</tr>
<tr>
<td>9</td>
<td>AtpB</td>
</tr>
<tr>
<td>10</td>
<td>beta-amylase</td>
</tr>
<tr>
<td>11</td>
<td>chloroplast protein</td>
</tr>
<tr>
<td>12</td>
<td>unnamed protein product</td>
</tr>
<tr>
<td>13</td>
<td>allergen Gly m Bd 28K</td>
</tr>
<tr>
<td>14</td>
<td>AtpA</td>
</tr>
<tr>
<td>15</td>
<td>seed maturation protein</td>
</tr>
<tr>
<td>16</td>
<td>HSP 70 kDa protein 1</td>
</tr>
<tr>
<td>17</td>
<td>protein disulfide isomerase</td>
</tr>
<tr>
<td>18</td>
<td>unknown protein</td>
</tr>
<tr>
<td>19</td>
<td>putative histone H2B</td>
</tr>
<tr>
<td>20</td>
<td>Enolase</td>
</tr>
<tr>
<td>21</td>
<td>alcohol dehydrogenase 1</td>
</tr>
</tbody>
</table>
Table 3

<table>
<thead>
<tr>
<th>Molecular Weight (kDa)</th>
<th>DSPC</th>
<th>SPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>621</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>493</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2 -