

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

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

16 **Abstract**

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

30 **Keywords:** Date seed proteins; proteomics; emulsifying properties

31

32 **1. Introduction**

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

46 The fruit of the date palm *Phoenix dactylifera L.* is one of the richest fruit-based
47 sources of protein. Date palm is one of the major fruit crops produced in dry and
48 semidry regions. It is an important commercial crop in different regions of the world
49 (Al-Yahyai & Manickavasagan, 2012) and is considered the third most important
50 palm species in the global agricultural industry after coconut and oil palms. The
51 seeds of the date fruit, which are a waste product from date processing, also contain
52 5–7% protein by weight (Aldhaferi et al., 2004), but very little is known about the
53 composition and the functional properties of these seed proteins. If it is possible to
54 extract the proteins from the seeds it might be useful as a source of protein for
55 human or animal nutrition. Robust methods for the extraction of proteins from date

56 seeds could facilitate the utilisation of date palm wastes such as seeds in the human
57 and animal diet.

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

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

72

73 **2. Materials & Methods**

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

75 **2.1 Preparation of Date Seed Protein Isolate**

76 Dates (i.e. the fruit of the date palm *Phoenix dactylifera L.*) were purchased from a
77 local supermarket in Edinburgh, United Kingdom. The dates were purchased at the
78 Tamr stage (complete maturity) and their variety was Deglet Nour that had been
79 grown in Tunisia. Seeds were removed from 40 kg of whole dates, washed in water
80 to remove any remaining date flesh and then air-dried for a week. The seed was
81 found to make up 10.3% (w/w) of the total mass of the date fruit on average. The
82 seeds were then further dried overnight at 40°C in a drying oven. The seeds were
83 milled using a hammer mill to a particle size that could pass through a 1–2 mm sieve
84 screen and then stored at –20°C until further preparation was required. The powder
85 obtained was identified as date palm seed powder (DPSP). The composition (w/w) of
86 the DPSP has been reported in our previous paper as protein, 5.64%, moisture,
87 5.39%, fat 8.14%, fibre 18.50%, ash 0.95%, carbohydrate 61.38% (Akasha,
88 Campbell & Euston, 2012).

89 Oil was extracted from DPSP using a Soxhlet apparatus. Fifteen gram samples of
90 dried DPSP were weighed into an extraction thimble (Fisher Scientific, UK) and
91 sealed with cotton wool. The thimble was inserted in a Soxhlet extraction flask and
92 extracted with boiling hexane (boiling point 68 °C) for 10 hours or until the solvent at
93 the sample chamber was colourless, indicating it was free from oil and that all the oil
94 had been extracted. The defatted DPSP was removed from the extraction thimble
95 and left to dry overnight to allow the hexane to evaporate. This defatted date seed
96 powder (DDSP) was kept at –20°C until processed further. The residual fat content
97 and protein content of the defatted powder were reported previously (Akasha,

98 Campbell & Euston, 2012) as 1.01% (w/w) and 6.13% (w/w) respectively. This
99 protein content is equivalent to a 100% yield of protein. The effect of the hexane
100 extraction step on the functionality of the proteins was not determined. However, it is
101 well known that the methods used to extract the proteins from the powder will also
102 affect the functionality so the additional effect of hexane extraction is likely to be
103 negligible.

104

105 **2.1.1 Protein Isolation**

106 Protein was extracted from the DDSP using a phenol/trichloroacetic acid (Ph/TCA)
107 extraction procedure based on the methods (with some modifications) proposed by
108 Gomez–Vidal et al., (2008) for olive and *Phoenix dactylifera L.* leaves respectively.
109 Ten grams of defatted DDSP was mixed with 30mL of ice–cold acetone, vortex
110 mixed and then centrifuged at 10,000rpm for 10 min at 4°C (Beckman Avanti J26-XP
111 centrifuge). The supernatant was decanted and discarded and the residual pellet
112 washed twice with ice–cold acetone and allowed to dry at room temperature. After
113 the pellet had dried it was ground to a fine powder using a pestle and mortar, rinsed
114 with 15% (w/v) TCA in acetone, vortex mixed and then centrifuged at 10,000 rpm for
115 10min at 4°C. The rinsing with TCA/acetone and centrifugation was repeated three
116 times. The pellet was then rinsed with cold 15% (w/v) TCA in water and centrifuged.
117 The rinsing with cold TCA and centrifugation was repeated three times. The pellet
118 was then rinsed with cold 80% (v/v) acetone followed by centrifugation, and this was
119 also repeated three times. The pellet was then air dried.

120

121 **2.1.2 Protein Purification**

122 To purify the protein two grams of the dry pellet was suspended in a mixture of 10mL
123 of Ph/Tris–buffer, pH 8.0 and 10 mL of dense SDS buffer (2%[w/v] SDS, 5%[w/v]
124 sucrose, 0.1M Tris-HCL, pH 8.0, 5% [v/v] β–mercaptoethanol). The mixture was
125 vortex mixed and the pellet was obtained by centrifugation at 10,000rpm for 10min at
126 4°C using a Beckman Avanti J26-XP centrifuge fitted with a JA25.50 rotor
127 (Beckman-Coulter, High Wycombe, UK). The pellet was resuspended in Ph/Tris–
128 buffer and dense SDS solution, and centrifuged again under the same conditions.
129 The pellets from both centrifugations were mixed and precipitated with five volumes
130 of cold 0.1M ammonium acetate in methanol, refrigerated at 4°C overnight and then
131 centrifuged at 10,000 rpm for 10min at 4°C. The pellet from this centrifugation was
132 then washed three times with cold methanol plus 0.1M ammonium acetate and
133 centrifuged as above followed by the same process with cold 80% (v/v) acetone. Half
134 a gram of the dried pellet was then mixed with 5 mL of cold aqueous 24% (w/v) TCA,
135 vortex mixed and left to precipitate on ice for 30min, followed by centrifugation at
136 13,000 rpm for 15 min at 4 °C. The pellet was washed with 2 mL of ice cold acetone,
137 incubated for 15 min on ice and then centrifuged at 13,000 rpm for 15 min at 4 °C.
138 The final pellet or date seed protein concentrate (DSPC) was air-dried in an oven at
139 30 °C overnight (16 hours) and stored at –20 °C until required for further analysis.

140

141 **2.1.3 Protein Content of DSPC**

142 The crude protein content of the extracted DSPC and DDSP was determined by
143 measurement of the nitrogen content using the Kjeldahl method (Lynch, Barbano &
144 Fleming, 1998).

145 The percent yield of protein from the date palm seed was determining by calculating
146 the protein recovered in the DSPC and comparing this to the maximum possible
147 protein recovery from the DDSP.

148

149 **2.2 SDS-PAGE Analysis of DSPC**

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

160

161

162 **2.3 Preparation of Protein for LC-MSMS**

163 Protein preparation was carried out using a method proposed by Le Bihan et al.,
164 (2011). Ten mg of DSPC was resuspended in 50 µL of distilled water (dH₂O),
165 followed by denaturation with 250 µL of 8M urea and dilution with 950 µL dH₂O prior
166 to trichloroacetic acid (TCA) precipitation with 310 µL of 100% TCA, 1250 µL
167 methanol and 625 µL chloroform. Samples were vortex-mixed and incubated (4°C,
168 10 min) before centrifugation (4,500g, 4 °C, 10 min). The top phase was removed
169 before adding 1 mL methanol. The sample was vortex mixed before centrifugation

170 (4,500xg, 4 °C, 10 min), the supernatants were removed and the solid sample
171 washed twice with 1 mL acetone, centrifuged at 10,000g, at 4 °C for 5 min and dried
172 under vacuum. Then, the sample was resuspended in 100 µL dH₂O.

173 Protein digestion was carried out using the method proposed by Le Bihan et al.,
174 (2011) on 20µL of protein extract. Briefly, samples were denatured in 8M urea,
175 reduced by incubating with dithiothreitol (DDT) prior to cysteine alkylation with
176 iodoacetamide and overnight digestion with 60 µg trypsin at room temperature. Four
177 µg of peptide samples were acidified with 1% formic acid before centrifugation and
178 cleaning using Stage tips (Thermo Scientific, Hemel Hempstead, UK). Finally, the
179 peptide samples were vacuum-dried and stored at -20 °C until further analysis.

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

187

188 **2.3.1 Identification and Quantification of the Peptides**

189 Multicharged (2+, 3+ and 4+) ion intensities were extracted from the LC-MS files and
190 the Mascot Version 2.4 software (Matrix Science Ltd, UK) was used to compare the
191 MSMS data against the NCBI protein database (13/03/2013; 11,961,441
192 sequences). Search parameters used were a maximum missed-cut value of 2,
193 variable oxidation (M), N-terminal protein acetylation and fixed
194 carbamidomethylation (C), precursor mass tolerance 7 ppm and MSMS tolerance

195 0.4 Da. A significance threshold (p) of <0.05 (MudPIT scoring) was set and a
196 minimum peptide cut off score of 20. Proteins identified and quantified with 2 or more
197 peptide sequences were retained.

198

199 **2.4 Emulsifying Properties**

200 Emulsifying activity index (EAI) and emulsion stability index (ESI) were determined
201 by a turbidimetric method according to Ogunwolu et al. (2009) with some
202 modifications. Four hundred and fifty milligrams of protein sample was dispersed in
203 45 mL of Mill-Q water. The protein solution was then mixed with 15 mL of sunflower
204 oil purchased from a local supermarket (Tesco Ltd, UK) and the pH was adjusted to
205 2, 4, 6, 8, 10 or 12 using 0.1M HCl or 0.1M NaOH. The mixture was homogenised
206 using an Ultra-turrax high speed homogenizer (IKA-Werke GmbH, Germany) for 1
207 min to make a protein-stabilised oil-in-water emulsion. Fifty μ L of the emulsion was
208 removed from the bottom of the container using a pipette and suspended in 5 mL of
209 0.1% (w/v) SDS solution. This was carried out immediately at 0 min and 10 min after
210 the homogenisation. Absorbance of the diluted emulsions was measured at 500nm
211 using a UV/Vis spectrophotometer (Model Genesys 6, Thermo Electron Corporation,
212 USA). The ability of the protein to form an emulsion (emulsifying activity index, EAI)
213 and the stability of the formed emulsion (emulsion stability index, ESI) were
214 calculated using the following formulae:

$$215 \quad \text{EAI (m}^2/\text{g)} = \frac{2 \times T \times A_0 \times \text{dilution factor}}{C \times \phi \times 1000}$$

$$216 \quad \text{ESI (min)} = \frac{A_0}{A_0 - A_{10}} \times \Delta t$$

217

218 Where, $T = 2.303$, A_0 = absorbance immediately after the homogenisation, dilution
219 factor = 100, C = the weight of protein per unit volume (g/mL), ϕ = the oil volumetric
220 fraction (0.25), A_{10} = absorbance after 10 min of the homogenisation, $\Delta t = 10$ min.
221 The emulsifying ability and emulsion stability was repeated in triplicate and the error
222 bars quoted as the standard deviation of the mean.

223

224 **3 Results & Discussion**

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

228

229 **3.1 Identification of the Date Palm Seed Protein Isolates by LC-MSMS**

230 Over three hundred proteins were detected in the DSPC sample by LC-MSMS. Not
231 all identifications were considered significant (see below). Protein identification was
232 achieved after the MSMS data were compared to known sequences on the NCBI
233 database using the Mascot Version 2.4 software (Matrix Science Ltd, UK). This
234 search resulted in 318 hits, each of which corresponding to a unique protein. The
235 protein list was screened to remove any contaminants (e.g. proteins that the
236 database only identified as being found in humans or animals). Since the preparation
237 method for the LC-MSMS requires digestion of the sample with trypsin, this protein,
238 corresponding to the hit number 1 (i.e. the most abundant protein) is ignored. A
239 second protein, keratin (hit number 59), an animal protein found in hair, nails and
240 skin, was also removed as this was considered to be a contaminant. To determine
241 how accurate the identification of the remaining proteins was we used two criteria,
242 the MOWSE score and the condition that the identification be based on at least two
243 peptides being matched to the predicted peptide map of the protein. MOWSE

244 (Molecular Weight Search) is a method that aids in identifying proteins based on
245 molecular weight of the peptides formed from proteolytic digestion of the protein
246 sample by allowing the probability of correct identification of the protein to be
247 calculated. The method was first developed by Pappin, Hojrup & Bleasby (1993).
248 This method calculates the probability that the peptide has been misidentified during
249 database searching, i.e. the identification is a random event. A low probability (P) of
250 misidentification is required for correct identification. Since it is more common to
251 express a more accurate identification as a higher number, the probability of
252 misidentification is converted to a MOWSE score using the formula,

253

$$254 \text{ MOWSE Score} = -10 \cdot \log_{10}(P) \quad (1)$$

255

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

$$268 \text{ MOWSE cut-off} = -10 \cdot \log_{10}\left(\frac{1}{20 \times 11961441}\right) = 83.7 \quad (2)$$

269

270

271 Therefore any protein match with a MOWSE score of 84 or greater will have less
272 than a 5% chance of being an incorrect identification. The first 111 hits were
273 considered to have been successfully identified since they all displayed a MOWSE
274 score of 85 or greater and therefore can be considered to be found in date palm
275 (*Phoenix dactylifera L.*) seed. However, on closer inspection not all of these have
276 been identified as a particular protein, with some being labelled unknown proteins,
277 and some hypothetical (identified from gene sequences) but which are nonetheless
278 in the NCBI database. Other proteins failed the second criterion that more than one
279 peptide is used in the identification. Once these proteins had been removed along
280 with contaminants, 90 unique proteins were identified. These 90 most abundant
281 proteins were classified into twelve different groups according to their functions using
282 the categories described by Bevan et al (1998). The different functional group
283 classifications and percentages found in the DSPC are show in Table 1.

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

288 The twenty most abundant proteins are listed in Table 2. Data listed in Table 2
289 include the hit number (HN), protein description, molecular weight search score
290 (MOWSE score), protein molecular weight (MW) and number of peptide matches
291 compared to total number of peptides produced. The hit number is a rough indicator
292 of protein abundance in the sample, with a higher hit number indicating a more

293 abundant protein. A discussion of the function of the twenty most abundant proteins
294 follows according to the functional category they belong to.

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

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

313 Ribulose-1,5-bisphosphate carboxylase (RuBisCo) (hit no. 8) is one of the most
314 abundant proteins on Earth being found in all green plants The biological function of
315 RuBisCO is to catalyze two reactions: the carboxylation of D-ribulose 1,5-
316 bisphosphate, the primary event in carbon dioxide fixation and the oxidative
317 fragmentation of the pentose substrate in the photorespiration process. The potential

318 of Rubisco as a food protein has been hypothesized for many years (Douillard & de
319 Mathan, 1994). Recently the focus has been on the extraction of rubisco from the
320 leaves of green plants, and this has revealed that, depending on the extraction
321 method, rubisco powders with good functional properties can be made (Kamm,
322 Kamm, Scherze, Muschiolik & Binbrich, 2006).

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

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

339 The most abundant protein band occurred at 60kDa, with minor bands identified at
340 higher and lower molecular weights for the date seed protein isolate. Using
341 non-reducing conditions (data not shown) did not alter the protein band profile
342 significantly, suggesting that disulphide bonds were absent from these proteins. It

343 was previously reported that albumins of oil palm seeds did not display disulphide
344 bonds (Morcillo et al., 1997). Khoshroo et al. (2011) reported similar results based on
345 an analysis of seed protein from twelve varieties of date palm (Bazmani sefid (Bw.Ji),
346 Mahminai, Gordial, Kharok, Almehtari, Mordar sang, Kaluteh, Halilehi, Bazmani sefid
347 (Bw.Ba), Mazafati, Khorbak syah, Khosh kang) grown in different Iranian regions.
348 The researchers found one heavily stained band at around 65kDa and minor bands
349 ranging from 12 to 369 kDa. Bouaziz et al. (2008) found three similar prominent
350 protein bands in date seeds of Allig and Deglet Nour varieties at 32, 60 and 70KDa',
351 The differences in protein profile between our results and the previous work (Bouaziz
352 et al., 2008) could be explained by a number of factors. The extraction process used
353 in the other studies differs from ours and this may lead to differential extraction of
354 proteins. Variation between the seed storage proteins is expected within different
355 varieties of the same species. In particular, extensive genetic polymorphism of seed
356 proteins is observed both within the same genotype and among genotypes of the
357 same species. This genetic polymorphism may occur through the presence of
358 multigene families within the same species, or through post-translational
359 glycosylation of proteins or proteolytic action on the proteins (Miernyk and Hajduch,
360 2011). Glycosylation, in particular, will lead to several proteins with the same amino
361 acid sequence but differing molecular weight due to the presence of one or more
362 sugar chains of variable length and position. Finally, the protein composition of the
363 seed varies during the embryo development process, with the major storage protein
364 not appearing until 3 months after fertilization. Thus, the level of maturity of the date
365 fruit will also influence the protein profile found in the seed. This may partly explain
366 the differences in molecular weight profile for the seeds proteins found in our study
367 and those of Bouaziz et al. (2008), Ehsanpour et al. (2010). Purification and

368 characterization of storage proteins in oil palm embryo (the same family as the date
369 palm) has been studied by Morcillo et al, (1997). They identified the major storage
370 proteins as being 2S and 7S globulins. The 2S proteins were made up of two
371 polypeptides (one acidic and one basic) of 22 kDa and 19 kDa molecular weight
372 respectively. The 7S proteins were the major fraction identified using SDS-PAGE.
373 These were shown to be a heterogeneous group of polypeptides of molecular weight
374 between 45 and 65 kDa with no disulphide bonds. They were also found in the form
375 of oligomers with molecular weights of 156 and 201 kDa.

376 For comparison purpose a soy protein isolate sample was also run on an SDS PAGE
377 gel. Six intense, detectable bands were observed (lane C, Figure 1), located at
378 approximately 535, 64, 50, 36, 22 and 16 kDa respectively. These bands might be
379 identified with basic polypeptides of glycinin which have an accepted molecular
380 weight range from 16-22KDa, acidic polypeptides of glycinin with molecular weight
381 range 34-36KDa, β -subunit (40-50KDa) and α -subunit (64KDa) (Roesch & Corredig,
382 2005). The high molecular weight band at 535 KDa could correspond to oligomers of
383 glycinin.

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

392 et al., 2001). The functional properties of these two proteins will be discussed further
393 below.

394 *Functional category 11:* Dakhlaoui-Dkhil et al. (2013) report that 16.6% of the
395 identified proteins of date palm (*Phoenix dactylifera L.*) leaf are defence-related
396 proteins which include defence regulated proteins and resistance proteins, those
397 involving detoxification, stress responses, cell rescue and cell death (Bevan et al.,
398 1998). This compares to 7% of date seed proteins in this category found in this study
399 (Table 1). The protein from this category which is found in the 20 most abundant was
400 the seed biotin-containing protein (hit no. 7).

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

406 The NCBI database of proteins which was searched using Mascot during the
407 analysis of the proteomics results is the largest store of experimentally identified
408 biological macromolecular structures available. However, in this database there are
409 a large number of proteins that have uncharacterized functions. Unnamed or
410 hypothetical proteins are often those that have been identified based on genome
411 sequencing of an organism, but the protein for which the gene codes has not been
412 identified, named and characterised in the plant or animal. Dakhlaoui–Dkhil et al
413 (2013) reported that 29.4% of protein detected in date palm leaf was hypothetical
414 protein, not dissimilar from the 22% detected here (Table 1).

415

416 **3.2 Emulsifying Activity and Emulsion Stability of Date Seed Proteins** 417 **Compared to SPI and WPC**

418 For DSPC to be exploited as a food ingredient it must show comparable functional
419 properties to other food proteins. The emulsifying properties (emulsifying ability (as
420 emulsifying activity index (EAI) and emulsion stability as emulsion stability index
421 (ESI)) were compared to soy protein isolate (SPI) and bovine whey protein
422 concentrate (WPC) in Figures 2 and 3. SPI is a common plant protein emulsifier, and
423 WPC is a highly functional animal protein emulsifier in formulated foods (Euston &
424 Hirst, 2000). The emulsifying properties were tested over a range of pH. At all pH
425 values the EAI of WPC was significantly greater than that of SPI which was in turn
426 significantly greater than that of DSPC (Figure 2). All three samples showed a
427 minimum in EAI at pH 4-5. This minimum occurred at the isoelectric point for both
428 WPC at pH 4.8 (Demetriades, Coupland & McClements, 1997) and soy proteins
429 between pH 4.7-5.0 (Golubovic, van Hateren, Ottens, Witkamp, van der Wielen,
430 2005). The same trends as seen for EAI between the three samples were not
431 observed with the emulsion stability. In Figure 3 the ESI proved very similar for all
432 three protein samples across the whole pH range. Furthermore, WPC emulsion ESI
433 was slightly less than for DSPC at most pH values.

434 The size of emulsion droplets is a major factor in the stability of the emulsion, with
435 larger droplets proving less stable than smaller droplets. Therefore, a correlation
436 might be expected between EAI and ESI since the EAI is an indirect measurement of
437 the droplet size. When the ESI is plotted against EAI a linear relationship between
438 ESI and EAI for all three protein samples is observed as expected, i.e. a larger EAI
439 (smaller particles size) resulted in more stable droplets (Supplementary Figure 1).
440 Differences in the relationship between EAI and ESI are observed between the three

441 protein samples. EAI values were similar for both DSPC and SPI, however the slope
442 of the EAI vs ESI graph was greater for the DSPC emulsions than for the SPI
443 emulsions, suggesting that for a given droplet size the SPI emulsions were less
444 stable. For WPC emulsions the EAI was high compared to DSPC and SPI
445 emulsions, but the ESI was lower for a given EAI, although the correlation was still
446 linear with a slope very similar to that for the SPI.

447 Soy bean protein emulsifying functionality has been widely studied (Utsumi,
448 Katsumura, & Mori, 1997). Soy proteins are predominantly glycinin and β -conglycinin
449 (70% of the total protein) and these two proteins determine the emulsifying
450 properties. The DSPC was shown above to contain high levels of glycinin and β -
451 conglycinin so we would expect these proteins to play a large part in the emulsifying
452 behaviour of DSPC.

453

454 The quaternary structure of both glycinin and β -conglycinin is complex. In the plant
455 seed, glycinin is found as a hexamer (molecular weight in the range 300–380 kDa),
456 and is made up of combinations of 5 distinct subunits (Staswick, Hermodson,
457 Nielsen, 1984). Glycinin hexamers can form trimers (7S) or monomers (3S) by
458 dissociation at different pH and ionic strength combinations (Peng, Quass, Dayton &
459 Allen 1984). β -conglycinin also forms oligomers comprised of three polypeptide
460 chains (α , α' and β) with overall molecular weight in the range 150–200 kDa (Thanh
461 & Shibasaki, 1979). The subunit composition of β -conglycinin is also variable. Soy
462 proteins have been found to form adsorbed layers 30–40 nm thin at the surface of oil
463 droplets (Keerati-u-rai & Corredig, 2010). Whey proteins such as β -lactoglobulin, on
464 the other hand, form adsorbed layers that are only 4–6 nm thick (Atkinson,
465 Dickinson, Horne & Richardson, 1995). The conclusion that can be drawn is that soy

466 proteins adsorb as aggregates (oligomers) rather than individual proteins unlike β -
467 lactoglobulin. Maruyama et al. (2004) found that the subunit composition of the
468 hexameric glycinin affects the emulsifying properties. Since the glycinin subunit
469 composition is variable (Staswick, Hermodson, Nielsen, 1984) the emulsifying ability
470 of soy proteins may vary. The subunit composition of β -conglycinin also affects
471 emulsifying ability (Utsumi, Matsumura & Mori, 1997). The α subunit has been shown
472 to be the best emulsifier followed by α' and then β (Utsumi, Matsumura & Mori,
473 1997). In addition, β -conglycinin is a better emulsifier than glycinin, due to its ability
474 to adsorb more rapidly at the emulsion droplet surface and to spread more
475 extensively at the interface (Utsumi, Matsumura & Mori, 1997; Bernard, Grandison &
476 Lewis, 2001; Molina, Papadapoulou & Ledward, 2001). Clearly, the relative
477 proportion of glycinin and β -conglycinin and their subunit composition affected the
478 emulsifying properties of the DSPC and SPI powders, and this could explain the
479 differences in emulsifying properties between the DSPC and SPI. Distinct differences
480 in the protein molecular weight profile between the DSPC and SPI was observed in
481 the SDS-PAGE results with a greater proportion of high molecular weight protein
482 fractions seen in the DSPC (Table 3 and Figure 1).

483 We can speculate as to why DSPC emulsions are more stable than SPI emulsions of
484 the same EAI (Supplementary Figure 1) by considering what is already known about
485 the emulsifying properties of aggregated proteins. It has been observed previously
486 (Euston & Hirst, 2000) that aggregated proteins are often poorer emulsifiers than
487 non-aggregated proteins. However, the aggregated proteins emulsions displayed a
488 greater stability under certain conditions. The explanation given for this was that the
489 aggregates display a greater conformational stability than the native proteins, and
490 were unable to unfold and spread rapidly to stabilise the oil droplet surface of small

491 droplets, which leads to larger droplets (lower EAI). On the other hand, because the
492 proteins are in the form of large aggregates the adsorbed protein layer around the
493 emulsions droplets is very dense and occupies a large volume, and leads to an
494 increased emulsion droplet stability. Euston & Hirst (2000) proposed that two
495 mechanisms were responsible for this increased emulsion stability. The density of
496 the emulsion droplet is increased by the presence of the dense aggregated protein
497 layer and this will reduce their creaming velocity and increase stability to creaming
498 (Euston & Hirst, 2000). In addition, the extensive aggregated protein adsorbed layer
499 is likely to increase the steric stabilising effect of the protein layer, thus reducing the
500 likelihood of coalescence (Euston & Hirst, 2000). We have seen the presence of
501 large oligomers of proteins in our DSPC, and in SPI (Table 3 & Figure 1) and this
502 may explain the higher stability of DSPC emulsions over SPI and WPC emulsions.
503 WPC proteins do form oligomers, but these are only loosely associated (Iametti,
504 Scaglioni, Mazzini, Vecchio & Bonomi, 1998) and easily break up under
505 emulsification conditions so that only protein monomers adsorb and a thin
506 monomeric layer of protein is adsorbed to the emulsion droplet surface. This will
507 have a lower steric stabilising ability, and lower effect on droplet density than the
508 aggregates found in DSPC (and SPI).

509 The DSPC sample contains 32% non-protein which is almost certainly complex
510 carbohydrate. We have carried out unpublished studies using various enzymes to
511 aid the extraction of the protein. These suggest that there are high proportions of
512 mannans, beta-glucans, xylans and cellulose present in the seeds, and that the seed
513 proteins are more closely associated with the glucans and cellulose. Sekhar and
514 DeMason (1988) have found that 75% of the protein in date palm seeds is found in
515 the cotyledon parenchyma cells (part of the embryo), whilst only 17% is found in the

516 endosperm, where the mannans are found. We would therefore expect the date
517 seed proteins to be associated with glucans, xylans and cellulose rather than
518 mannans. The presence of these polysaccharides in the protein sample will influence
519 the functional properties of the proteins. Recently, Bouaziz et al. (2013) have studied
520 the functional properties of fibro-protein complexes from date seed and have
521 demonstrated that they have potential as emulsifiers in food applications.

522

523 **4 Conclusions**

524 In this study we have extracted protein from the seeds of the date fruit and
525 characterised them using proteomic analysis. LC–MSMS revealed a large number of
526 proteins in the date seed protein sample. Of the 90 proteins identified with high
527 confidence (MOWSE score above 84) the majority of these proteins (70% by
528 number) have metabolic functions in the seed and seedling, whilst of the remainder
529 15% (by number) are storage proteins such as 11S and 7S globulin (glycinin and β -
530 conglycinin) (Table 1). The emulsifying properties of DSPC were determined and it
531 was found to have a comparable to SPI.

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

541 also foaming and gelation, will need to be characterised over a wider range of
542 conditions that are relevant to food systems.

543 **Acknowledgements**

544 IA acknowledges the receipt of a PhD Scholarship from the Libyan government.

545 Mass spectrometry was carried out at the Kinetic Parameter Facility, University of
546 Edinburgh, UK.

547 **5 References**

548 Akasha, I.A., Campbell, L. & Euston, S.R. (2012). Extraction and characterisation of
549 protein fraction from date palm fruit seeds. *World Academy of Science, Engineering*
550 *and Technology*, **6**, 10-25.

551 Aldhaferi, A., Alhadrami, G., Aboalnaga, N., Wasfi, I. & Elridi, M. (2004). Chemical
552 composition of date pits and reproductive hormonal status of rats fed date pits. *Food*
553 *Chemistry*, **86**, 93–97.

554 Al-Yahyai & Manickavasagan (2012). An overview of date palm production. In A.
555 Manickavasagan, M.M. Essa, & Sukumar, E. (eds), *Dates: production, processing,*
556 *food, and medicinal values* (pp 3-12). Boca Raton: CRC Press.

557 Andreou, A. & Feussner, I. (2009). Lipoxygenases - Structure and reaction
558 mechanism. *Phytochemistry*, **70**, 1504-1510.

559 Atkinson, P.J. Dickinson, E., Horne, D.S. & Richardson, R.M. (1995). Neutron
560 reflectivity of adsorbed β -casein and β -lactoglobulin at the air/water interface.
561 *Journal of the Chemical Society, Faraday Transactions*, **91**, 2847–2854.

562 Bernard, E.C., Grandison A.S. & Lewis, M.J. (2001). Emulsifying properties of soy
563 protein isolate fractions obtained by isoelectric precipitation. *Journal of the Science*
564 *of Food and Agriculture*, **81**, 759–63.

565 Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R.,
566 Dirske, W., Van Staveren, M., Stiekema, W. et al. (1998). Analysis of 1.9 Mb of
567 contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature*, **391**, 485–
568 488.

569 Bouaziz, M., Besbes, S., Blecker, C., Wathelet, B., Deroanne, C., & Hamadi, A.
570 (2008). Protein and amino acid profiles of Tunisian Deglet Nour and Allig date palm
571 fruit seeds. *Fruits*, **63**, 37–43.

572 Dakhlaoui-Dkhil, S.S., Coquet, L. Cosette, P., Elkahoui, S., Song, P.C.T., Vaudry, D.,
573 Hannachi, A.S., Trifi, N., Limam, F. & Jonenne, T. (2013). The Date palm (*Phoenix*
574 *dactylifera* L.) leaf proteome: identification of a gender biomarker to screen male
575 parents. *Plant Omics Journal*, **6**, 18–23.

576 Demetriades, K. Coupland J.N. & McClements, D.J. (1997). Physical properties of
577 whey protein stabilized emulsions as related to pH and NaCl. *Journal of Food*
578 *Science*, **62**, 342–347.

579 Douillard, R. & de Mathan, O. (1994). Leaf protein for food use: potential of Rubisco.
580 In B.J.F. Hudson (Ed), *New and Developing Sources of Proteins* (pp 307– 342). New
581 York: Springer.Ehsanpour, A.A., Shojaie, B. & Rostami, F. (2010). Characterization
582 of seed storage protein patterns of four Iranian pistachios using SDS–PAGE. *Natural*
583 *Science*, **2**, 737–740.

584 Euston, S.R. & Hirst, R.L. (2000). Comparison of the concentration-dependent
585 emulsifying properties of protein products containing aggregated and non-
586 aggregated milk protein, *International Dairy Journal*, **9**, 693–701.

587 Golubovic, M., van Hateren, S.H., Ottens, M., Witkamp, G.J. & van der Wielen, L.A.
588 (2005). Novel method for the production of pure glycinin from soybeans. *Journal of*
589 *Agricultural and Food Chemistry*, **53**, 5265–5269.

590 Gomez-Vidal, S., Tena, M., Lopez-Liorca, L.V. & Salinas, J. (2008). Protein
591 extraction from *Phoenix dactylifera* L. leaves, a recalcitrant material, for two-
592 dimensional electrophoresis. *Electrophoresis*, **29**, 448-456.

593 Hamada, J.S., Hashim, I.B. & Sharif, F.A. (2002). Preliminary analysis and potential
594 uses of date pits in foods. *Food chemistry*, **76**, 135–137.

595 Iametti, S., Scaglioni, L., Mazzini, S., Vecchio, G. & Bonomi, F. (1998). Structural
596 features and reversible association of different quaternary structures of β -
597 lactoglobulin. *Journal of Agricultural and Food Chemistry*, **46**, 2159–2166.

598 Kamm, B., Kamm, M., Scherze, I., Muschiolik, G. & Binbrich, U. (2006). Biobased
599 polymers by chemical valorization of biomass components. In C. Vasile & G.E.
600 Zaikov (Eds) *New Trends in Natural and Synthetic Polymer Science* (pp. 15-24).
601 New York: Nova Science Publishers.

602 Keerati-u-rai, M. & Corrediq, M. (2010). Heat-Induced changes occurring in
603 oil/water emulsions stabilized by soy glycinin and β -conglycinin. *Journal of*
604 *Agricultural and Food Chemistry*, **58**, 9171–9180.

605 Khoshroo, S.M.R., Khavarinjad, R., Fahimi, H., & Mohammadi, Z. N. (2011). Seed
606 storage protein electrophoretic profiles in some Iranian date palm (*Phoenix*
607 *dactylifera* L.) cultivars. *African Journal of Biotechnology*, **10**, 17793–17804.

608 Kong, X., Li, X., Wang, H., Hua, Y. & Huang, Y. (2008). Effect of lipoxygenase
609 activity in defatted soybean flour on the gelling properties of soybean protein isolate.
610 *Food Chemistry*, **106**, 1093–1099.

611 Le Bihan, T., Martin, S.F., Chirside, E.S., van Ooijen, G., Barrios-Llerena, M.E.,
612 O'Neill, J.S., Shliaha, P.V., Kerr, L.E. & Millar, A.J. (2011). Shotgun proteomic
613 analysis of the unicellular alga *Ostreococcus tauri*. *Journal of Proteomics*, **74**, 2060-
614 2070.

615 Lynch, J. M., Barbano, D. M. & Fleming, J. R. (1998). Indirect and direct
616 determination of the casein content of milk by Kjeldahl nitrogen analysis:
617 collaborative study. *Journal of AOAC International*, **81**, 763–774.

618 Martin, S., Munagapati, V.S., Salvo-Chirnside, E., Kerr, L. & Le Bihan, T. (2012).
619 Proteome turnover in the green alga *Ostreococcus tauri* by time course 15N
620 metabolic labeling mass spectrometry. *Journal of Proteome Research*, **11**, 476-486.

621 Maruyama, N., Prak, K., Motoyama, S., Choi, S.K., Yagasaki, K., Ishimoto, M.,
622 Miernyk, J.A. & Hajduch, M. (2011). Seed proteomics. *Journal of Proteomics*, **74**,
623 389–400.

624 Molina, E., Papadapoulou, A. & Ledward, D.A. (2001). Emulsifying properties of
625 high-pressure-treated soy protein isolate and 7S and 11S globulins. *Food*
626 *Hydrocolloids*, **15**, 263–9.

627 Morcillo, F., Aberlenc-Bertossi, F. & Trouslot, P., Hamon, S. & Duval, Y. (1997).
628 Characterization of 2S and 7S storage proteins in embryos of oil palm. *Plant*
629 *Science*, **122**, 141-151.

630 Nakabayashi, T., Takakusagi, Y, Iwabata, K. & Sakaguchi, K. (2011). Foam
631 fractionation of protein: Correlation of protein adsorption onto bubbles with a pH-
632 induced conformational transition. *Analytical Biochemistry*, **419**, 173–179.

633 Ogunwolu, S.O., Henshaw, F.O., Mock, H.P. & Santos, A. (2009). Functional
634 properties of protein concentrates and isolates produced from cashew (*Anacardium*
635 *occidentale* L.) nut. *Food Chemistry*, **115**, 852–858.

636 Pappin D.J., Hojrup P. & Bleasby A.J. (1993). Rapid identification of proteins by
637 peptide-mass fingerprinting. *Current Biology*, **3**, 327-332.

638 Peng, I.C., Quass, D.W., Dayton, W.R. & Allen, C.E. (1984). The physicochemical
639 and functional properties of soybean 11S globulin a review. *Cereal Chemistry*, **61**,
640 480–490.

641 Pimentel, D. & Pimentel, M. (2003). Sustainability of meat-based and plant-based
642 diets and the environment. *American Journal of Clinical Nutrition*, **78**, 660S-663S.

643 Roesch, R. & Corredig, M. (2005). Heat-induced soy-whey proteins interactions:
644 formation of soluble and insoluble protein complexes. *Journal of Agricultural and*
645 *Food Chemistry*, **53**, 3476–3482.

646 Schutyser, M.A.I. & van der Goot, A.J. (2011). The potential of dry fractionation
647 processes for sustainable plant protein production. *Trends in Food Science &*
648 *Technology*, **22**, 154–164.

649 Sham, A. & Aly, M.A.M. (2102). Bioinformatics based comparative analysis of
650 omega-3 fatty acids in desert plants and their role in stress resistance and tolerance.
651 *International Journal of Plant Research*, **2**, 80-89.

652 Smith, A.M., Zeeman, S.C. & Smith, S.M. (2005). Starch degradation. *Annual*
653 *Review of Plant Biology*, **56**, 73-98.

654 Staswick, P.E., Hermodson, M.A. & Nielsen, N.C. (1984). Identification of the
655 cysteines which link the acidic and basic components of the glycinin subunits.
656 *Journal of Biological Chemistry*, **259**, 13431–13435.

657 Thanh, V.H. & Shibasaki, K. (1979). Major proteins of soybean seeds. Reversible
658 and irreversible dissociation of β -conglycinin. *Journal of Agricultural and Food*
659 *Chemistry*, **27**, 805–809.

660 Utsumi, S., Matsumura, Y. & Mori, T. (1997). Structure-function relationships of soy
661 proteins. In S. Damodaran & A. Paraf (Eds), *Food Proteins and Their Applications*
662 (pp. 257-292). New York: Marcel Dekker.

663 Utsumi, S., Katsube, T., Ishige, T. & Takaiwa F. (1997). Molecular design of soybean
664 glycinins with enhanced food qualities and development of crops producing such
665 glycinins. *Advances in Experimental and Medical Biology*, **415**, 1–15.

666 Xu, C., Garret, W.M., Sullivan, J., Caperna, T.J. & Natarajan, S. (2006). Separation
667 and identification of soybean leaf proteins by two-dimensional gel electrophoresis
668 and mass spectrometry. *Phytochemistry*, **67**, 2431–2440.

669

670 **Table Legends**

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

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

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

681

682 **Figure Legends**

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

687 Figure 2 – Emulsifying activity index (EAI) as a function of pH for date seed protein
688 concentrate (DSPC), soy protein isolate (SPI) and whey protein concentrate (WPC).
689 Error bars are \pm one standard deviation of the mean.

690 Figure 3 – Emulsion stability index (ESI) as a function of pH for date seed protein
691 concentrate (DSPC), soy protein isolate (SPI) and whey protein concentrate (WPC).
692 Error bars are \pm one standard deviation of the mean.

693

694 **Table 1**

Protein Functional Category	Proteins in this Category (%)
Metabolism/sugars and polysaccharides/amino acids/Nucleotides/Lipid	15
Energy/ ATP synthase/ Glycolysis/ Electrontransport/ Gluconeogenesis/ Photosynthesis/ Pentose phosphate	33
Cell growth/ division	0
Transcription/ mRNA	1
Protein synthesis/ Translation factors	8
Protein destination and storage/ Storage protein	10
Transporters/ Transport ATPases	3
Intracellular traffic	0
Cell structure	0
Signal transduction	1
Stress responses/Disease/defence/pathogenesis-related protein	7
Unclear classification	22

695

696

697 **Table 2**

698

HN	Proteins description	MOWSE score	MW (Da)	Peptides matched
2	Glycinin	2436	54927	86/99
3	alpha subunit of beta conglycinin	1624	63184	52/74
4	Lipoxygenase	1001	97490	40/48
5	Sucrose-binding protein	878	60884	34/42
6	unnamed protein product	855	22972	24/25
7	Seed biotin-containing protein	654	67894	20/24
8	ribulose-1,5-bisphosphate carboxy.	652	53056	35/48
9	AtpB	406	51944	14/16
10	beta-amylase	399	56378	15/19
11	chloroplast protein	347	26530	8/9
12	unnamed protein product	341	47117	11/13
13	allergen Gly m Bd 28K	328	52780	9/9
14	AtpA	324	54044	10/12
15	seed maturation protein	312	17907	11/13
16	HSP 70 kDa protein 1	312	71420	10/13
17	protein disulfide isomerase	308	58963	12/19
18	unknown protein	299	43082	13/16
19	putative histone H2B	284	14338	2/2
20	Enolase	232	48127	7/11
21	alcohol dehydrogenase 1	127	20101	3/5

699

700

701
702
703

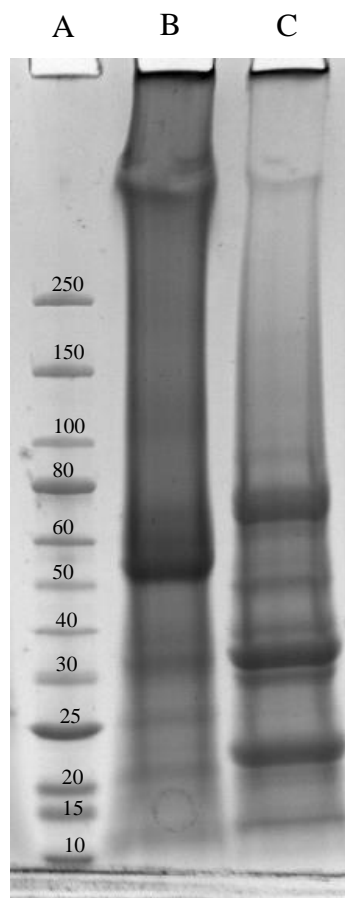
Table 3

Molecular Weight (kDa)	
DSPC	SPI
621	535
493	113
150	82
83	72
72	64
62	54
60	50
34	41
32	36
27	35
25	30
20	22
18	20
	16
	13

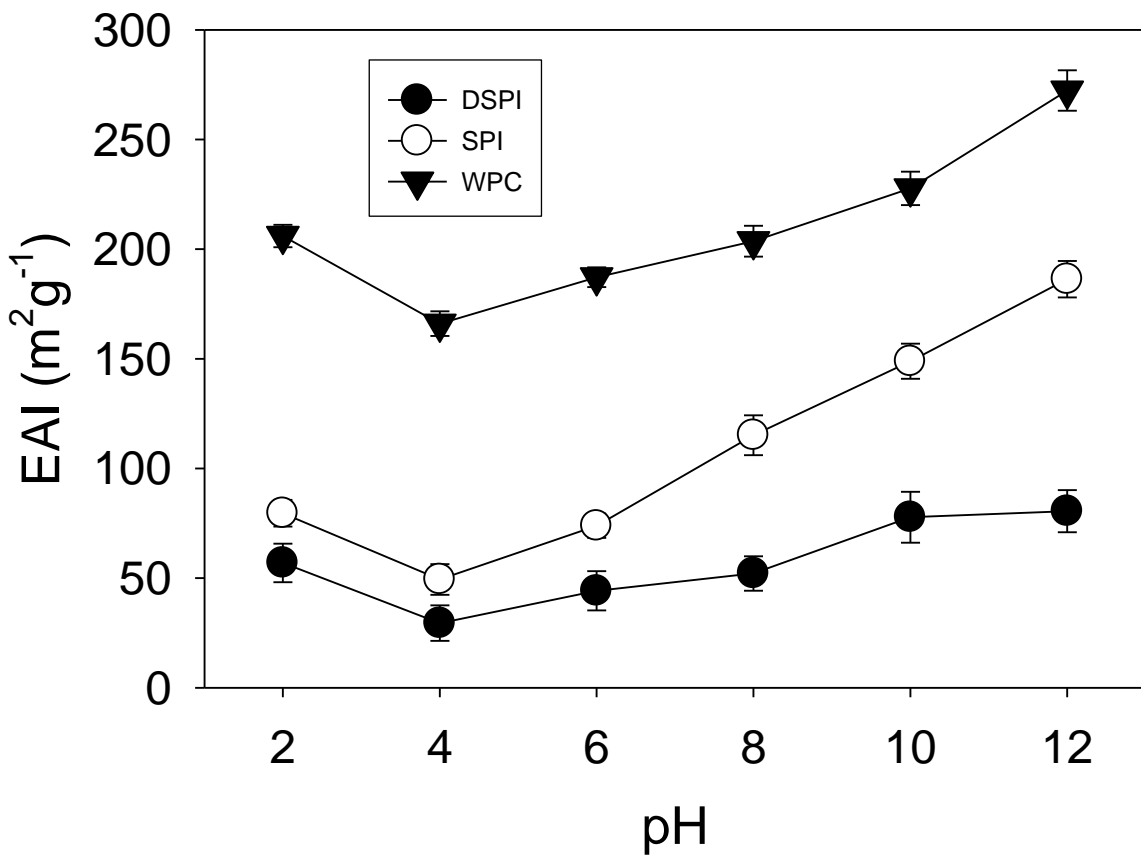
704
705

706

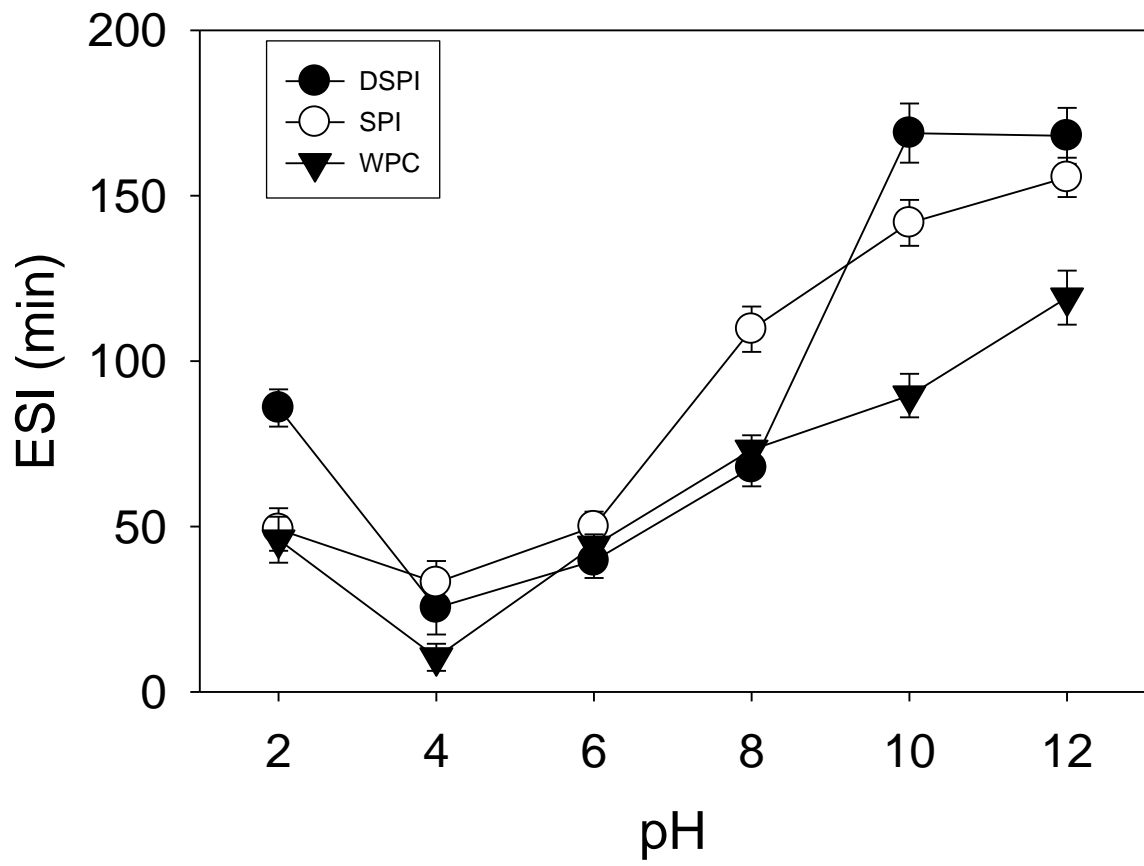
707



708 Figure 1



709
710
711 Figure 2 -
712
713



714

715

716

717 Figure 3-