



HAL
open science

Identification and characterization of the main peptides in pea protein isolates using ultra high-performance liquid chromatography coupled with mass spectrometry and bioinformatics tools

Audrey Cosson, Lydie Oliveira Correia, Nicolas Descamps, Anne Saint-Eve,
Isabelle Souchon

► To cite this version:

Audrey Cosson, Lydie Oliveira Correia, Nicolas Descamps, Anne Saint-Eve, Isabelle Souchon. Identification and characterization of the main peptides in pea protein isolates using ultra high-performance liquid chromatography coupled with mass spectrometry and bioinformatics tools. *Food Chemistry*, 2022, 367, pp.130747. 10.1016/j.foodchem.2021.130747 . hal-03517046

HAL Id: hal-03517046

<https://hal.inrae.fr/hal-03517046>

Submitted on 2 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title:** Identification and characterization of the main peptides in pea protein isolates using ultra high-
2 performance liquid chromatography coupled with mass spectrometry and bioinformatics tools

3 **Authors:** Audrey Cosson^{ab}, Lydie Oliveira Correia^c, Nicolas Descamps^b, Anne Saint-Eve^a, Isabelle
4 Souchon^{d*}

5 Affiliations

6 ^aUniv Paris Saclay, UMR SayFood, AgroParisTech, INRAE, F-78850 Thiverval Grignon, France.

7 ^bRoquette Frères, 10 rue haute loge, F-62136, Lestrem, France

8 ^cUniv Paris Saclay, INRAE, AgroParisTech, Micalis Inst,PAPPSO, F-78350 Jouy En Josas, France

9 ^dAvignon Univ, UMR SQPOV, INRAE, F-84000 Avignon, France

10 ***Corresponding author:** Isabelle Souchon, Phone: +33 (0)4 32 72 24 89, Address: UMR408 SQPOV
11 - Sécurité et Qualité des Produits d'Origine Végétale, Domaine Saint Paul, 228, route de
12 l'Aérodrome, Site Agroparc - CS 40509, 84914 Avignon Cedex 9, FRANCE

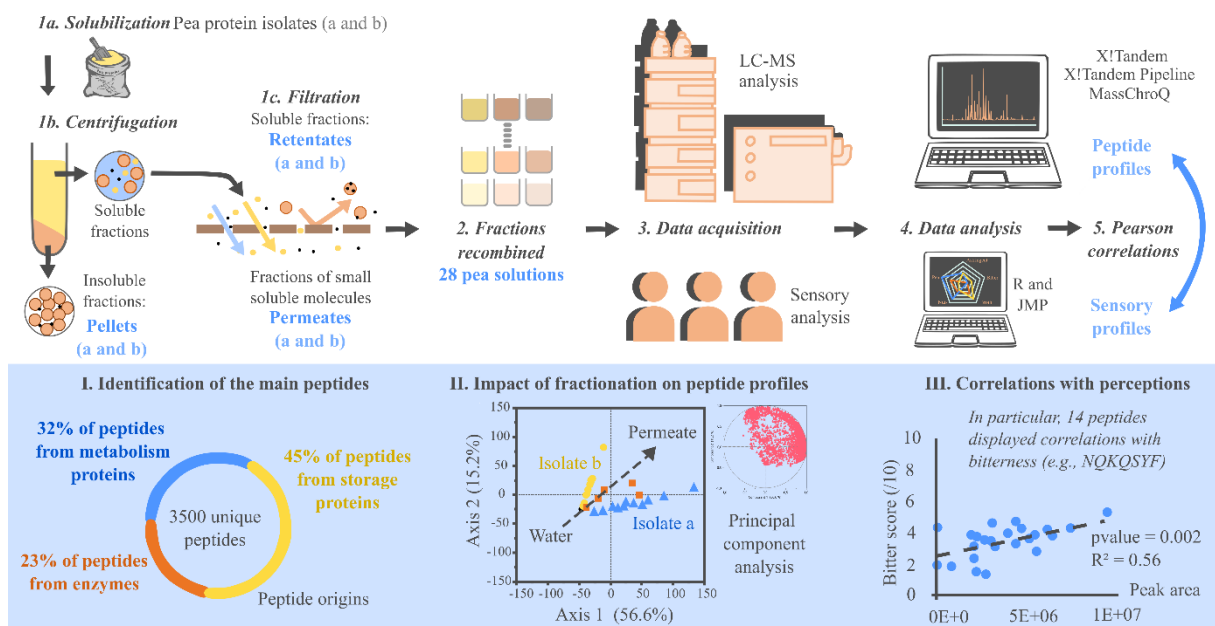
13 **Email:** isabelle.souchon@inrae.fr

14 **Email addresses for co-authors:**

15 audrey.cosson@inrae.fr ; lydie.oliveira-correia@inrae.fr ; nicolas.descamps@roquette.com ;

16 anne.saint-eve@inrae.fr

17 **Graphical abstract:**



18

19 **Highlights**

- 20 • 3,005 unique peptides identified in pea protein solutions
- 21 • 45% of peptides came from seed storage proteins, mainly vicilins
- 22 • 11 peptides displayed sequence homology with known antioxidants
- 23 • 1,640 peptides were associated with high broth scores, perhaps reflecting umami
- 24 • 14 peptides appeared to influence the perception of bitterness

25

26 **Abstract:**

27 Pea protein isolates are a source of high-quality plant proteins. However, from a sensory perspective,
28 they are usually described as having strong beany and bitter notes, which arise from a complex
29 mixture of volatiles, phytochemicals, and peptides. The aim of this study was to identify the main
30 peptides in isolates and examine their correlations with sensory perceptions. Thus, 28 solutions
31 containing different mixtures of pea protein fractions were assessed. Any peptides present were
32 identified and characterized using ultra high-performance liquid chromatography-mass spectrometry.
33 There were a total of 3,005 unique peptides representing various protein families; 1,640 and 275
34 peptides were correlated with broth and bitter attributes, respectively. In particular, 14 peptides with
35 short sequences (< 8 residues) were correlated with bitterness. These results show how key peptides in
36 isolates may cause sensory perceptions.

37 **Keywords:** pulse, peptidomics, bitter, beany, sensory, correlations

38

39 **1. Introduction**

40 A major recent challenge in the agrifood industry is developing new protein sources to compensate for
41 the anticipated future paucity of traditional animal proteins. Consequently, both the industry and
42 consumers are focusing their attention on plant proteins. Plant protein isolates, such as those derived
43 from peas (*Pisum sativum* L.), are often used to create foods because of their functional properties,
44 protein content, sustainable production, and relatively low cost (Davis et al., 2010). However, plant
45 proteins, and especially isolate fractions from raw plant matter, have some drawbacks from a sensory

46 point of view (e.g., their color, smell, and taste). It is necessary to better understand the sensory issues
47 associated with plant proteins if we wish to develop plant protein-based foods that will be attractive to
48 consumers.

49 Research on the perception of pea-based products has largely focused on the role of volatile aroma
50 compounds in creating sensations of beaniness (Bi et al., 2020) and of phenolics/saponins in creating
51 sensations of bitterness and astringency (Heng et al., 2006). However, it is important to carry out more
52 detailed compositional analyses to clarify how foods are sensorily perceived.

53 Pea protein isolates are mainly composed of globulins, which are the main storage proteins in seeds.
54 Globulins consist of two fractions that are characterized by their ultracentrifugation sedimentation
55 coefficients: 7S (20%–40%) and 11S (20–30%). The 7S fraction is composed of vicilins and
56 convicilins. The 11S fraction is composed of legumines (Creveiu-Gabriel, 1999). During protein
57 isolate extraction (notably during temperature and pH changes), proteins may be naturally hydrolysed
58 into numerous peptides of different sizes (Li & Aluko, 2010; Sirtori et al., 2012). Several structural
59 changes result because of the exposure of hydrophobic sites normally found in the protein's core
60 (Daher et al., 2020). Although such peptides remain little studied, they could potentially have
61 properties that might serve to improve the sensory properties of plant-based products.

62 Indeed, specific protein fragments may elicit various sensory perceptions (e.g., sweet, bitter, umami,
63 sour, or salty notes). Sourness and saltiness could result from the presence of charged terminal groups
64 and/or charged side chains (Temussi, 2012). Other perceptions (sweetness, umami, and bitterness)
65 could be explained by the presence of different peptide families. For example, certain small peptides
66 (5–8 residues in size) can activate the TAS2R bitter taste receptors in the mouth (Aubes-Dufau et al.,
67 1995; Maehashi & Huang, 2009). These peptides tend to be hydrophobic with proline- and leucine-
68 rich side chains, especially at their C-terminals (Kim et al., 2008). They can have quite an impact: for
69 example, 0.25 mM of a peptide (VVYPWTQRF) solution derived from bovine hemoglobin elicits the
70 same sensation of bitterness as 0.073 mM of quinine sulfate or 21 mM of caffeine (Aubes-Dufau et al.,
71 1995). With regards to sweetness, there are no known natural peptides that result in sweet notes.

72 However, semi-synthetic peptides, such as aspartame (the methyl ester of the aspartic
73 acid/phenylalanine dipeptide) and neotame (a secondary amine of 3,3-dimethylbutanal and aspartame),

74 can activate T1R2/T1R3 sweet taste receptors. In the case of umami, the umami heterodimer
75 (T1R1/T1R3) has ligands with multiple binding sites, and thus the receptor displays low specificity
76 and can respond to a chemically diverse range of umami molecules. More than 50 peptides (such as
77 KGDEESLA) appear to elicit umami, but their specific functional roles remain unclear. Research on
78 the relevant receptors has suggested that such peptides might directly lead to the sensation of umami.
79 However, it is also possible that umami is a consequence of partial hydrolysis, which leads to sizeable
80 concentrations of Asp or Glu (Temussi, 2012; Wang et al., 2020).

81 Several experimental approaches have been used to study the sensory properties of protein fractions.
82 The most common strategy to examine how specific compounds affect the sensory characteristics of
83 products, using a combination of fractionation and omission tests (Engel et al., 2002; Toelstede &
84 Hofmann, 2008). However, peptidomics techniques are increasingly used thanks to advances being
85 made in modern mass spectrometry and bioinformatics. These tools are ideally suited for carrying out
86 comprehensive peptide analysis, especially when such analyses exploit the massive quantities of
87 information currently available in genomic and transcriptomic databases. In peptidomics, different
88 solvents and techniques are used in the fractionation, separation, and analysis of peptides (Gao et al.,
89 2019; Salger et al., 2019). In such work, liquid chromatography-mass spectrometry is the most widely
90 used analytical method. Fragmentation spectra obtained from samples are compared with theoretically
91 expected spectra for peptide reference sequences. Sample peptides are thus assigned to the proteins
92 that contain their sequences. Several bioinformatics tools have been developed to automate these
93 operations, such as COMET (Eng et al., 2013) or X!tandempipeline (Langella et al., 2017).

94 Information about peptide properties can be found in databases such as BIOPEP (Iwaniak et al., 2016).
95 Recently, Daher et al. (2020) demonstrated that peptidomics could be a valuable tool for evaluating the
96 bitterness of protein isolates.

97 Thus, the aim of this study was to identify the main oligopeptides and polypeptides (5–40 amino acids
98 long) found in pea protein isolates and to characterize their sensory properties. To this end, we used
99 pea protein solutions and an experimental design previously employed by Cosson et al. (2021). The
100 peptide profiles of the solutions were determined using ultra high-performance liquid chromatography
101 coupled with mass spectrometry (UHPLC-MS/MS). The resulting peptides were identified, and both

102 their physicochemical properties and their antioxidative properties were characterized. Then, we
103 examined the impact of our fraction-based formulation strategy on peptide profile. Finally, the
104 relationship between peptide profiles and the sensory properties of solutions (as determined in Cosson
105 et al., 2021) was explored, with a particular focus on perceived bitterness.

106

107 **2. Materials and methods**

108 **2.1. Solution preparation**

109 To obtain a pea protein isolate, different unit fractionation steps (precipitation, centrifugation,
110 membrane separation) followed by heat treatments are implemented. These processing steps influence
111 sensory characteristics as well as functional property of products (Gharibzahedi & Smith, 2021;
112 Roland *et al.*, 2017). Thus, two commercial pea protein isolates were used in this study. Six fractions
113 were obtained from two pea protein isolates (protein content = nitrogen [N] content x 6.25; 83% dry
114 matter) as explained in Cosson et al. (2021): permeates a and b; retentates a and b; and pellets a and b.
115 The main elements of fractionation are recalled here. The isolates were dispersed in tap water to obtain
116 a suspension (4% (w/w) dry matter content) and maintained under agitation for 12 h at 3 °C. Then, it
117 was centrifuged (Jouan Kr4i and a Sorvall Lynx 4000 [Thermo Scientific, Waltham, US]; 6000 g, 10
118 min, 4 °C) and the supernatant was manually separated from the pellet. The pellet was then diluted
119 with tap water (12.35% (w/w) dry matter content). Then, a tangential filtration module (TIA, Bollene,
120 France) was used with two ST-3B-1812 PES Synder membranes (46-mil spacer; 10-kDa MWCO) and
121 a high-pressure diaphragm pump (Wanner Hydra-Cell G10, Wanner International Ltd, Church
122 Crookham, UK). Throughout filtration, the retentate was at 13 °C, the inlet pressure (P1) was at 1.5
123 bar, the outlet retentate pressure (P2) was at 1 bar, and the mean transmembrane pressure ($(P1 + P2)/2$)
124 was at 1.25 bar. First, ultrafiltration was used to obtain 10 L of permeate; then, diafiltration was
125 performed to partially wash the retentate with one diavolume.
126 Then, these six fractions were combined in various ways to formulate the 28 unique and different
127 solutions of the mixture design described in Cosson et al., 2021 (see Supplementary Table 1). Among
128 these solutions Refa (respectively Refb) correspond to the solutions of pea protein isolates a (resp. b)
129 at 4% (w/w). This process was carried out at 4°C in 50 mL glass flasks, which were stored at -20°C.

130 Work was performed on different groups of compound types rather than on a single compound type.
131 Each fraction was associated with a main compound type: insoluble proteins in the case of the pellets;
132 soluble compounds (e.g., volatiles, peptides, and phenolics) in the case of the permeates; and soluble
133 proteins interacting with volatiles in the case of the retentates. This approach made it possible to
134 formulate a diversity of pea-protein-based solutions to obtain continuous responses and build reliable
135 statistical models. Solutions were chosen to represent a broad spectrum of combinations while also
136 remaining realistic in terms of protein concentrations (0–4.25%). The different steps to obtain the
137 solutions and the solutions analysis are described in the Supplementary Figure 1.

138

139 **2.2. Overall characterization of the solutions**

140 For each fraction, protein content was determined via the Kjeldahl method (N content x 6.25). Dry
141 matter content (% w/w) and ash content were determined by a certified external laboratory (SAS
142 IMPROVE, Amiens, France) via drying and calcination (prepASH[®] 219 analysis system). Conductivity
143 at 20°C was measured with a calibrated conductivity probe (InPro 7108-25/65-VP 3.B, M300
144 transmitters; Mettler Toledo, Switzerland), and pH was measured at 20°C (InPro 4801i/SG/120;
145 Mettler Toledo, Switzerland).

146 Hydrophobicity index values were measured as per Kato and Nakai (1980). The reaction between the
147 8-anilinonaphthalene-1-sulphonic acid probe (ANS) and hydrophobic amino acids (Alanine, Valine,
148 Leucine, Isoleucine, Methionine, Phenylalanine, Tryptophan, Proline) leads to the formation of a
149 fluorescent complex, which is measured by spectrofluorometry. Each solution was diluted with
150 phosphate buffer (0.053 M Na₂HPO₄-2H₂O, 0.067 M KH₂PO₄, pH = 7.0) to establish five
151 concentration levels between 0.002% and 0.032% (wt). Then, 20 µL of an 8-anilinonaphthalene-1-
152 sulphonic acid probe (ANS; Sigma Aldrich) was added to 4 mL of each solution; the result was
153 thoroughly mixed for 15 min in the dark (concentration of 8 mM in the phosphate buffer). Signal
154 intensity was measured using a spectrofluorometer (Cary Eclipse; Agilent) with excitation
155 wavelengths of 380nm and emission wavelengths of 480nm. Protein surface hydrophobicity was then
156 calculated as the initial slope of relative fluorescent intensity in function of the protein concentration.
157 The relative fluorescent intensity was calculated as $(F-F_0)/F_0$ where F_0 is the fluorescent intensity

158 values of the ANS blanks (ANS solution made with buffer - without any proteins). F is the fluorescent
159 intensity values of the protein solutions. The slope of the relationship between protein concentration
160 (%) and fluorescent intensity was determined via linear regression analysis. Three replicates of the
161 analysis were performed.

162

163 **2.3. Peptide identification and relative quantification**

164 Before the UHPLC-MS/MS analysis, a sample pre-treatment procedure was applied adapted from
165 previous work (Guillot et al., 2016). Due to the nature of the samples, which are the result of several
166 fractionation steps, it was possible to simplify the sample preparation procedure as follows. Pea
167 solutions were centrifuged (15,000 g, 4°C, 15 min). The supernatants were filtered using a Vivaspin
168 centrifugal concentrator (20 mL, 10 kDa; Sigma Aldrich) run at 8,000 g (30 min, 4°C). The filtrates
169 were stored in the dark at -80°C prior to analysis.

170 MS was performed at the PAPPSO platform (MICALIS, INRAE, Jouy-en-Josas, France). An Orbitrap
171 Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) coupled to an UltiMate™
172 3000 RSLCnano System (Thermo Fisher Scientific) was used. Peptides were loaded into a precolumn
173 (Acclaim PepMap C18; 5 µm particle size, 5 mm length, 300 µm ID) at a rate of 20 µL/min and were
174 separated using a C18 column (Acclaim PepMap nanoViper; 2 µm particle size, 500 mm length, 75
175 µm ID) at a rate of 300 nl/min and measured over a total gradient length of 147 min with increasing
176 buffer B (80% acetonitrile [ACN] and 0.1% formic acid) from 1 to 60 % for 115 min. Buffer A was
177 0,1% formic acid in 98% water. The eluted peptides were distributed throughout the gradient showing
178 a good and an adequate peptide separation (Supplementary Figure 2). The eluted peptides were
179 analyzed online using the Orbitrap mass analyzer. The mass spectrometer was operated in data
180 dependent acquisition (DDA) and positive mode ionization was performed, employing a spray voltage
181 of 2.8 kV. Peptide ions were analyzed using a data-dependent method as follows: a full MS scan (m/z:
182 300–1,600; resolution: 120,000) was performed by the Orbitrap mass analyzer. Doubly and triply
183 charged peptides underwent MS/MS analysis (collision energy: 30%; resolution: 30,000; cycle time: 3
184 sec).

185 Peptide identification was performed with X!Tandem v. 2017.2.14 (Alanine) and X!Tandem Pipeline
186 (C++) v. 0.2.40 (Langella et al., 2017) using protein sequences for *Pisum sativum* L. The main peptide
187 identification parameters were the following: no cleavage specificity, variable methionine oxidation
188 state, and mass tolerance for parent and fragment ions of ± 10 ppm. Peptides were retained when the E-
189 value was ≤ 0.05 , and the presence of one peptide per parental protein was considered to enable
190 identification. Contaminant peptides were discarded following identification using a standard
191 proteomics contaminant database, and the false discovery rate was estimated using the reversed
192 protein database.

193 MassChroQ software (v. 2.2.17) was employed to perform alignment, XIC extraction, peak detection,
194 and quantification (Valot et al., 2011).

195 Fourteen pea solutions were analyzed using UHPLC-MS/MS: Refa, Refb, 100Pa, 100Pb, 100Ra,
196 100Rb, 50Ia-50W, 50Ib-50W, 50Pb-25Ib-25W, 25Pa-25Ra-13Ia-38W, 70Pb-30Ra, 40Ra-30Ib-30W,
197 50Ra-25Ia-25W, and 50Pb-50Rb. Among them, 100Ra, 100Rb, Refa, and Refb were performed in
198 duplicate to assess method repeatability.

199 To cut down on the analysis time, we hypothesized that the solutions' peptide concentrations could be
200 estimated from solution formulations, given that the solutions were mixtures of the fractions. We
201 found support for this hypothesis using a subset of six of the solutions (50Pb-25Ib-25W, 25Pa-25Ra-
202 13Ia-38W, 70Pb-30Ra, 40Ra-30Ib-30W, 50Ra-25Ia-25W, and 50Pb-50Rb). For the other thirteen
203 solutions (50Pa-Ia25-W25, 50Rb-50W, 40Pb-60W, 50Pb-50W, 40Pa-60Rb, 30Ia-70W, 60Ra-40W,
204 50Pa-25Ib-25W, 40Pa-60W, 50Pa-50Ra, 40Rb-30Ia-30W, 25Ib-75W, and 40Rb-30Ib-30W), peptide
205 composition was calculated based on the peptide composition of the fractions. A linear equation of the
206 following type was used:

$$207 \quad A_{\text{Recombined.Products}} = A_{\text{pellet.a}} \times C_{\text{pellet.a}} + A_{\text{retentate.a}} \times C_{\text{retentate.a}} + A_{\text{permeate.a}} \times C_{\text{permeate.a}} + A_{\text{pellet.b}} \times C_{\text{pellet.b}} +$$
$$208 \quad A_{\text{retentate.b}} \times C_{\text{retentate.b}} + A_{\text{permeate.b}} \times C_{\text{permeate.b}}$$

209 where A was the area of the peptides, and C was the relative quantity of each fraction.

210 Before the statistical analyses were performed, the data were processed. First, the areas of each
211 replicate were averaged (cleansing step). The areas of identical peptides with the same charges were
212 also summed. In this study, we chose to use all a peptide's isotopes in its quantification. This decision

213 was made for two reasons: a) the isotope distribution for a given peptide is discrete and depends
214 mainly on the presence of heavy isotopes and b) isotope composition can be treated as
215 "homogeneous." Consequently, using all the isotope peaks should improve the results because signal
216 variability should decline if multiple values are used. Missing data are always a modeling concern, so
217 we assumed that this approach would still yield a better approximation than comparing isotopes
218 separately. Second, certain peptides were removed (first filtering step): only peptides present in at least
219 two solutions were retained. Third, the peptide composition of 13 of the solutions was calculated as
220 described above (calculation step). Fourth, peptides with little variation in area were removed (second
221 filtering step): only peptides that varied at least 50% among the solutions were retained. Finally, to
222 remove any artefacts, null values were replaced by randomly selected values between $1+E04$ and
223 $1+E05$ (i.e., values corresponding to the detection threshold). The general workflow of the different
224 steps of peptidomics analysis is illustrated below (Fig. 1).

225

226 **2.4. Characterization of peptide properties**

227 Peptides were characterized based on nine physicochemical properties: length (number of amino
228 acids), the GRAVY index (the grand average of hydrophobicity), bulk (the average bulkiness of the
229 amino acids), the aliphatic index (relative volume occupied by aliphatic side chains), polarity (average
230 polarity of the amino acids), charge (overall net charge), relative basic nature (fraction of informative
231 positions that are occupied by Arg, His, or Lys), relative acidic nature (fraction of informative
232 positions that are occupied by Asp or Glu), and relative aromatic nature (fraction of informative
233 positions that are occupied by His, Phe, Trp, or Tyr). As in Proust et al. (2019), these properties were
234 computed using the aminoAcidProperties function of the R package "alakazam" v. 0.2.8 (Gupta et al.,
235 2015). Default settings were used for scaling and normalization. The bioactivity and sensory
236 properties of peptides were explored via comparisons with known bioactive and taste peptides listed in
237 the BIOPEP database (Iwaniak et al., 2016). Only peptides that were more than three amino acids long
238 were examined to avoid noise in the results. Finally, the perceptions of peptides were investigated by
239 looking at the sensory scores of the 28 solutions evaluated by Cosson et al. (2021).

240

241 **2.5. Statistical analysis**

242 Analyses were performed using R (R Core Team, 2019) and JMP (v. 13.1.0; SAS Institute Inc., Cary,
243 SC, USA). For the inferential analyses, $\alpha = 0.05$ was the threshold for statistical significance. To
244 visualize the intersections in the peptide sets among the six fractions and the two raw solutions, the
245 function “upset” in the package UpSetR was used (Conway et al., 2017). To visually explore
246 differences in peptide profiles among the 28 solutions, we carried out principal component analysis
247 (PCA, wide method) on a correlation matrix. To visualize the overall characteristics of the peptides,
248 we plotted the distributions of each physicochemical property (normalized distribution, kernel
249 density). Finally, to examine the relationships between the peptide data and the sensory data for the 28
250 solutions, we analyzed a correlation matrix (Pearson method).

251

252 **3. Results and discussion**

253 **3.1. Identification and characterization of the main peptides in pea protein isolate solutions**

254 *3.1.1. Identification of the peptides in the pea protein solutions*

255 After preliminary processing of the peptide data, 3,561 peptide ions (with different charges) and 3,005
256 unique peptides were identified. Mass ions varied in m/z (305–1395 m/z), charge (2–4), isotope
257 number (0–5), and area ($1.0E+04$ – $1.0E+10$, median = $1.4E+06$). The three most common peptides
258 were NPFIFK, FANAQPQQR, and NQKQSYF; they came from vicilins and provicilins. They likely
259 represent favored hydrolysis sites. In addition, 348 peptides with the following modifications were
260 identified: loss of an ammonia, usually via vicinal dehydration, ammonia rearrangement, and
261 rehydration via ammonia release, resulting in the loss of nitrogen without any gain in oxygen
262 (MOD:01160); oxygenation of an L-methionine residue to form a diastereomeric L-methionine
263 sulfoxide residue (MOD:00719); replacement of a residue amino or amino hydrogen with an acetyl
264 group (MOD:00408); and formation of a double bond via the removal of a water molecule from a
265 residue (MOD:00704) (Jupp et al., 2015).

266 The 3,005 peptides had origins in a wide range of proteins from three main groups (Fig. 2): storage
267 proteins (45%), enzymes (23%), and proteins derived from seed metabolism (32%). Within these

268 groups, only the proteins with the most peptides are illustrated. The others have been grouped
269 according to their functions. The majority of the peptides came from storage proteins and more
270 specifically, from vicilins (18%), convicilins (4%), and legumins (11%). This result is not surprising
271 since protein isolates are mainly composed of the latter three protein types (Crevieu-Gabriel, 1999).
272 It was interesting to note the presence of peptides from proteins associated with sensory off-notes.
273 There were large quantities of peptides from lipoxygenases (7%), which catalyze the degradation of
274 polyunsaturated fatty acids; the latter are thought to play an important role in the development of
275 undesirable off-flavors in pulses. The initial products of lipoxygenase activity are hydroperoxides,
276 which are further degraded into a wide range of compounds, including many that are responsible for
277 off-flavors, such as hexanal and n-pentylfuran (Roland et al., 2017). In addition, peptides from
278 aldehyde dehydrogenase were observed. This enzyme catalyzes the oxidation of aldehydes and so can
279 modify the composition of volatile compounds of pea protein solutions and so the sensations of
280 beaniness. Peptides from protein that catalyzes phenolic acids modifications were observed:
281 carotenoid cleavage dioxygenase, chalcone synthase, gibberellin dioxygenase, and isoflavone synthase
282 were present. Phenolics acids play also a role in the development of undesirable off-flavors in pulse
283 (bitter and astringent notes).

284 The peptides displaying modifications were generally associated with three types of proteins:
285 lipoxygenases (10% of modified peptides), histones (12% of modified peptides), and ribosomal
286 proteins (5% of modified peptides). Peptides from these protein types could probably more sensitive to
287 modifications during the pea processing.

288 Thus, these results show that a wide variety of peptides were identified. These peptides represent
289 proteins from different families, mainly seed storage proteins. Clarifying the origin of these peptides
290 also gives us information about the proteins present in the isolates, including which proteins may
291 cause sensory perceptions (e.g., the lipoxygenases).

292

293 *3.1.2. Physicochemical properties of the peptides in the pea protein solutions*

294 The peptides' physicochemical and antioxidative properties were characterized. The nine
295 physicochemical properties were chosen with a view to comprehensively describing pea peptide

296 diversity. The normalized distributions of the property values for the peptides are in Figure 3. The
297 peptides were mostly polar and hydrophilic. The mean GRAVY index value was around -0.5, which is
298 also the overall mean value for the 20 standard amino acids (i.e., -0.49; Kyte & Doolittle, 1982). The
299 median net charge was close to zero. Mean bulk was around 15 Å, which is also the overall mean for
300 the 20 standard amino acids (i.e., 15.4 Å; Zimmerman et al., 1968). In terms of amino acid
301 composition, the peptides had more aliphatic amino acids (Ala, Val, Leu, and Ile) than aromatic amino
302 acids (His, Phe, Trp, and Tyr) or acidic amino acids (Asp and Glu). Finally, average length was 10
303 residues, although this observation should be interpreted with caution given the specificities of the
304 analytical pipeline. Indeed, the upper limit on length was defined by the purification process and, more
305 specifically, by the ultrafiltration steps; the lower limit on length (no peptides < 6 residues were
306 detected) was a direct consequence of the chosen MS detection range (300–1,600 m/z). Thus, these
307 results show that the main peptides in the pea protein isolates varied greatly in their physicochemical
308 properties; however, when the overall averages were obtained, they generally corresponded to the
309 averages for the 20 standard amino acids.

310 In addition, some of the identified peptides matched with antioxidant peptides observed in pea and
311 included in BIOPEP database (see Supplementary Table 2). Bioactive peptides are usually 2–20 amino
312 acids long and have molecular masses of less than 6 kDa (Sarmadi & Ismail, 2010; Sun et al., 2004).
313 Here, eleven peptides had sequences that were homologous with those of known antioxidant peptides
314 (BIOPEP database) previously identified in pea-protein-based solutions (Iwaniak et al., 2016): ADGF;
315 ADVFNPR; ELLI; FVPH; HLHP; KFPE; LPILR; SAEHGSLH; SGAF; YLKT; and YVGD. These
316 peptides contained many copies of phenylalanine, an amino acid known to mediate antioxidant activity
317 (Sarmadi & Ismail, 2010). They came from different proteins—storage proteins such as legumins;
318 enzymes such as seed linoleate 9S-lipoxygenase-3; and metabolic proteins such as transporters. These
319 results highlight that the diversity of peptides present in pea-protein-based solutions may have
320 nutritional benefits and could be used to enhance the value of plant-based foods. Peptide composition
321 should be studied further from a nutritional point of view.

322

323 **3.2. Impact of fractionation and recombination on peptide profiles**

324 *3.2.1. Impact of fractionation on peptide profiles*

325 This study adopted an original approach: the decision was made to work with fractions instead of
326 compounds because i) we had no a priori hypothesis on which compounds would be linked to
327 perceptions and ii) from a sensory point of view working with all compounds (e.g. with molecular
328 fractionation and omission tests) can be very long and difficult. We broke down the pea protein
329 isolates into six fractions (two pellets, two retentates, and two permeates), which were then
330 recombined to form different solutions using a mixture design. Before studying the recombined
331 solutions, we studied the impact of fractionation on peptide profile composition.

332 The overall characteristics of the six fractions and the two raw solutions are presented in Table 1. The
333 number of peptides per fraction was linearly correlated with the sum of the areas of the peptides
334 (Pearson method; $R^2 = 0.83$). The permeates (100Pa and 100Pb) contained the greatest number of
335 peptides, followed by the raw solutions (Refa and Refb). The pellets (50Ia-50w and 50Ib-50W) had
336 the lowest number of peptides. Solutions from pea protein isolate b (Refb, retentate 100Rb, permeate
337 100Pb and pellet 50Ib-50W) had fewer identified peptides overall than solutions from pea protein
338 isolate a (Refa, retentate 100Ra, permeate 100Pa and pellet 50Ia-50W). These differences could come
339 from the processing of the two commercial products. A perspective to this work could be to study and
340 identify the step (or steps) of the processing that generates these differences in peptide composition.

341 To visualize the intersections in peptide sets among the six fractions and the two raw solutions (Rfa
342 and Refb), an UpSet plot was used (Fig. 4). An higher number of peptides were observed in solutions
343 from Refa (permeate 100Pa, retentate 100Ra, refa and then pellet 50Ia-50W) than in the respective
344 products from Refb (permeate 100Pb, retentate 100Rb, refb and pellet b). However, there were more
345 peptides in the raw solution from Refb than in the raw solution from Refa. Thus, the two pea protein
346 isolates were not impacted in the same way by the pea processing: it would appear that more specific
347 peptides were “lost” from isolate b.

348 To understand the effect of fractionation on the peptide profiles of the fractions, we examined the
349 relationship between the fractions’ physical characteristics (Table 1) and the sum of the areas of the
350 peptides. There was not a significant correlation between peptide area and either dry matter content,

351 protein content, ash content, pH, or surface hydrophobicity. However, there was a significant linear
352 correlation with conductivity (Pearson method; $R^2 = 0.84$). Peptides (e.g., salts, which drive
353 conductivity) are rather soluble and small in size. During the centrifugation step, they must have
354 mostly gone into the supernatant, and then, during the filtration step, they must have passed into the
355 permeate. Protein content and conductivity were slightly higher in isolate a than in isolate b. It can be
356 assumed that these properties explain the higher peptide concentrations in the fractions from batch a.
357 The peptides' physicochemical properties showed similar normalized distributions across fractions
358 (Fig. 3). The only notable differences occurred in charge between the pellets (50Ib-50W and 50Ia-
359 50W) and the permeates (100Pb and 100Pb) and in the polarity between the raw solutions (Refa versus
360 Refb). The peptides in the permeate solutions varied slightly more in charge. The peptides in the raw
361 solution Refb were slightly more polar. Therefore, the fractionation process did not lead to peptide
362 profiles that differed in physicochemical properties. As for the eleven peptides with sequences
363 homologous with those of known antioxidant peptides, they occurred across the range of solutions.
364 They were, however, most common in the raw solutions (Refa and Refb).

365 *3.2.2. Peptide profiles of the recombined solutions*

366 A mixture design was used to create a suite of solutions by combining the pea protein fractions in
367 different ways. To validate this methodology, the peptide profiles of six of the recombined solutions
368 (50Pb-25Ib-25W, 25Pa-25Ra-13Ia-38W, 70Pb-30Ra, 40Ra-30Ib-30W, 50Ra-25Ia-25W, and 50Pb-
369 50Rb) were determined; the results were compared with the peptide profiles that were calculated using
370 the fraction-based approach. Considering the number of values to be compared (3,561 peptides x 6
371 solutions), we did not contrast the individual values of the recombined solutions but rather the
372 distributions of the differences between their measured and calculated values. These distributions were
373 compared to the distributions for replicates of the experimental replicate solutions (100Ra, 100Rb,
374 Refa and Refb). The quartiles were calculated excluding any null values. The quartiles for repeated
375 solutions were: 1st quartile— $4.39E+05$, median— $1.18E+06$, 3rd quartile— $2.39E+06$, and
376 maximum— $6.11E+08$. The differences between the quartiles for the measured versus calculated
377 values were as follows: 1st quartile— $7.19E+05$, median— $1.65E+06$, 3rd quartile— $4.22E+06$, and
378 maximum— $1.47E+09$. The overall distributions were similar between replicate solutions and the data

379 for the recombined solutions. The quartiles were slightly lower in the case of the former, but the orders
380 of magnitude were similar. In the case of some peptides, there were significant differences between the
381 measured and calculated values for the recombined solutions. However, these peptides were among
382 those with the largest areas, and the relative differences were therefore small. In conclusion, it
383 appeared that peptide profiles could be reliably estimated for the recombined solutions using fraction-
384 based calculations.

385 PCA was used to visually assess the main differences among the recombined solutions (Fig. 5). The
386 solutions were well distributed along axes F1 and F2, which accounted for 71.8% of the variance.
387 Thus, the maps based on the first two axes seemed to provide a good-quality projection of the initial
388 multidimensional table, even though some information might have remained hidden in the subsequent
389 axes. The data for the areas of the 3,561 peptides were clustered within one half of the correlation
390 circles and were thus clearly correlated. Overall, peptide concentrations increased from water solution
391 (X, lower left) to the more permeate-based solutions (100 Pb and 100Pa; X upper/right). The solutions
392 formulated with fractions from batch a (Fig. 5: in green) and the solutions formulated with fractions
393 from batch b (Fig. 5: in blue) stand out clearly. Batch a variability is mainly found on axis 1, and batch
394 b variability is mainly found on axis 2. Solutions formulated with fractions from both batch a and b are
395 in the middle (Fig. 5: in orange). Regardless of the batch, permeates (100Pa and 100Pb) had the
396 highest peptide concentrations. The pellets (50Ia-50W and 50Ib-50W) had the lowest peptide
397 concentrations. The raw solutions (Refa and Refb) had intermediate peptide concentrations.
398 Consequently, with this experimental design, we have managed to create two different ranges of
399 peptide concentrations. The peptides in fraction mixtures (batch a and batch b) allowed us to explore
400 any interactions.

401 **3.3. Identifying factors influencing perceived bitterness**

402 *3.3.1. Sensory properties of the recombined solutions*

403 The 6 different fractions were combined in various ways to formulate 28 pea protein solutions (see
404 Supplementary Table 1). These solutions had been used by Cosson *et al.*, in a previous sensory study
405 to obtain greater insight into the origin of perceived beaniness (expressed via the following attributes:
406 almond, broth, cereals, nuts, pea, and potato), bitterness, and astringency in pea-protein-based foods.

407 Cosson *et al.*, found that the attributes contributing to perceived beaniness were mainly influenced by
408 the retentate and permeate fractions, likely because of their levels of volatiles, which were indirectly
409 reflected by hexanal levels. Perceived astringency was mainly influenced by the retentate and pellet
410 fractions, while perceived bitterness was largely driven by the retentate fraction. Bitterness and
411 astringency were associated with the levels of phenolics, which were indirectly reflected by caffeic
412 acid content. However, this previous study concluded that a more detailed analysis of solution
413 composition (i.e., beyond hexanal and caffeic acid levels) would be needed to uncover the more
414 precise origins of these sensory perceptions (Cosson *et al.*, 2021).

415 Drawing upon the results of this previous study, the peptide profiles of the pea protein solutions were
416 examined in tandem with the sensory profile data. The correlations between peptide areas and sensory
417 scores were evaluated (Pearson method). Each sensory attribute (score out of 10) was correlated with
418 the areas of several peptides (p-value < 0.05). Correlations were most common for the broth attribute
419 (1,640 out of 3,561 peptides), followed by the salty attribute (1,277 out of 3,561 peptides). In contrast,
420 correlations with other attributes were significantly less frequent: bitter—275 peptides; astringent—
421 173 peptides, mouthfeel—410 peptides, pea—440 peptides, potato—246 peptides, almond—80
422 peptides, nuts—135 peptides, and cereals—214 peptides. We also compared the peptides from this
423 study with the sensory peptides listed in the BIOPEP database (Iwaniak *et al.*, 2016), but there were no
424 noteworthy findings.

425 Perceived saltiness can arise from the presence of peptides with charged terminals and/or charged side
426 chains (Temussi, 2012). However, the phenomenon can also have an indirect cause: NaCl and peptides
427 (small soluble molecules) are likely distributed in a similar way in pea protein fractions. Furthermore,
428 a higher salt concentration can also change how protein hydrolysis or peptide fractionation plays out,
429 resulting in different peptide concentrations (Cheison & Kulozik, 2017).

430 Broth notes may be perceived when peptides have activated T1R1/T1R3 umami receptors. Indeed,
431 umami is often described as a meaty, broth-like, or savory taste and can participate in perceived
432 brothiness (Lioe *et al.*, 2010). Such peptides are between 2 and 11 residues long. For example, Glu-
433 Gly-Ser-Glu-Ala-Pro-Asp-Gly-Ser-Ser-Arg was found to elicit the sensation of umami during the
434 consumption of peanut hydrolysate (Su *et al.*, 2012). However, the idea that some peptides activate

435 umami receptors is controversial because when such peptides are synthesized, they do not always elicit
436 umami (Maehashi et al., 1999). Another explanation could be that the sensation of umami is a
437 consequence of the peptides' partial hydrolysis, which results in sizeable concentrations of Asp or Glu
438 (Temussi, 2012; Wang et al., 2020). Considering the number of peptides associated with the broth note
439 (36% of the peptides), several mechanisms are likely at work. In any case, peptides appear to play a
440 major role in the construction of brothiness.

441

442 *3.3.2. Relationships between peptide presence and bitterness*

443 Although bitterness is correlated with a much smaller number of peptides, it is important to discuss
444 this sensation as well. Indeed, the bitterness of pea-protein-based foods is a major off-note in these
445 products (Roland et al., 2017). Here, among the 275 peptides correlated with bitterness, 106 were
446 exclusively correlated with bitterness. Many peptides have the ability to activate bitter receptors
447 (Aubes-Dufau et al., 1995). Based on past research, such peptides are between 5 and 8 residues long
448 (Aubes-Dufau et al., 1995; Maehashi & Huang, 2009). Here, however, only 14 of the peptides
449 associated with bitterness were less than 8 residues in length: SRNPIY, KRHGEW, NLQNYR,
450 SNKFGKF, NQKQSYF, YLKGLKF, YQKSTEL, APHWNIN, AQPLQRE, ISLNKIRL,
451 NQKQSYFA, ANAQPLQR, NAQPLQRE, and EVLSWSFH. The peptides KRHGEW, NQKQSYF,
452 NAQPLQRE, and AQPLQRE are particularly noteworthy because they were positively correlated
453 with the bitterest solutions. In contrast, the peptides YLKGLKF, YQKSTEL, and EVLSWSFH were
454 negatively correlated with bitterness. The correlations are shown on the Supplementary Figure 3.
455 Using BitterX software (Huang et al., 2016), it was found that these eight peptides were highly likely
456 to activate bitter receptors (either TA2R7 or T2R40; probability: 88–79%). However, to confirm that
457 these peptides contribute to perceived bitterness in the mouth, it would be necessary to study their
458 effects on bitter receptors *in vitro* or to have a sensory panel evaluate them in solution. It would also
459 be useful to assess their concentrations relative to their perception thresholds. For example, Toelstede
460 and Hofmann (2008) found that 12 peptides eliciting bitter sensations had recognition thresholds
461 between 0.05 and 6.0 mmol/L.

462 In addition, while other groups of peptides correlated with sensory perceptions had largely overlapping
463 characteristics, bitterness-related peptides displayed certain differences (Fig. 3), such as lower
464 GRAVY values (i.e., are more hydrophilic on average), lower aliphatic values (i.e., have smaller
465 relative volumes on average), and higher polarity (i.e., are more polar on average). These results do
466 not concur with the results of previous studies, which have shown that such peptides are mainly
467 hydrophobic (Kim et al., 2008). In conclusion, these peptides may affect sensations of bitterness in the
468 mouth by activating bitter receptors (i.e., the peptides displaying positive correlations), blocking bitter
469 receptors (i.e., the peptides displaying negative correlations), or interacting with other molecules that
470 do either (i.e., the peptides displaying positive or negative correlations).

471

472 **4. Conclusion**

473 In this study, we identified and characterized the main oligopeptides and polypeptides (5–40 amino
474 acids long) found in pea protein solutions. We had four main findings. First, we identified a wide
475 variety of peptides representing a range of protein families, mainly those containing seed storage
476 proteins but also those containing proteins that can play a role in sensory perceptions, such as
477 lipoxygenases. Second, these peptides were mostly polar and hydrophilic, and our fraction-based
478 formulation strategy did not affect their overall physicochemical properties. Third, eleven peptides had
479 sequences homologous with those of known antioxidant peptides. These results indicate that the
480 variety of peptides present in pea protein solutions can have nutritional benefits. Fourth, most of the
481 peptides in the pea protein solutions were correlated with sensory attributes. In particular, many
482 peptides were correlated with salty and broth attributes, perhaps expressing the relationship of some
483 peptides to umami. A lower but still significant number of peptides displayed a correlation with
484 bitterness. These results highlight the mechanistic importance of these molecules in sensory
485 perceptions in the mouth. Taken together, these results suggest that a better understanding of the
486 peptide composition of plant protein isolates could help us address related sensory issues and develop
487 plant-protein-based foods whose taste appeals more to consumers.

488 **5. CRediT authorship contribution statement**

489 Audrey Cosson: Methodology, Investigation, Formal analysis, Writing - Original Draft. Lydie Oliveira
490 Correia: Resources, Investigation. Nicolas Descamps: Funding acquisition. Anne Saint-Eve:
491 Methodology, Supervision, Writing - review & editing. Isabelle Souchon: Conceptualization,
492 Supervision, Writing - review & editing.

493

494 **6. Acknowledgments**

495 This work was funded by Roquette (Lestrem, France), the French National Research and Technology
496 Agency (ANRT-CIFRE 2017/0815), AgroParisTech (Paris, France), and the French National Research
497 Institute for Agriculture, Food, and Environment (INRAE). The authors thank the PAPPSO platform
498 (<http://pappso.inra.fr>), which is supported by INRAE (<http://www.inrae.fr>); the Ile-de-France regional
499 council (<https://www.iledefrance.fr/education-recherche>); IBiSA (<https://www.ibisa.net>); and CNRS
500 (<http://www.cnrs.fr>), which granted access to its mass spectrometry facilities. The authors also thank
501 David Forest for providing technical support.

502

503 **7. References**

504 Aubes-Dufau, I., Seris, J.-L., & Combes, D. (1995). Production of peptic hemoglobin hydrolyzates:
505 Bitterness demonstration and characterization. *Journal of Agricultural and Food Chemistry*, 43(8),
506 1982–1988. <https://doi.org/10.1021/jf00056a005>
507 Bi, S., Xu, X., Luo, D., Lao, F., Pang, X., Shen, Q., Hu, X., & Wu, J. (2020). Characterization of key
508 aroma compounds in raw and roasted peas (*Pisum sativum* L.) by application of instrumental and
509 sensory techniques. *Journal of Agricultural and Food Chemistry*, 68(9), 2718–2727.
510 <https://doi.org/10.1021/acs.jafc.9b07711>
511 Cheison, S. C., & Kulozik, U. (2017). Impact of the environmental conditions and substrate pre-
512 treatment on whey protein hydrolysis: A review. *Critical Reviews in Food Science and Nutrition*,
513 57(2), 418–453. <https://doi.org/10.1080/10408398.2014.959115>

514 Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
515 intersecting sets and their properties. *Bioinformatics*, 33(18), 2938–2940.
516 <https://doi.org/10.1093/bioinformatics/btx364>

517 Cosson, A., Blumenthal, D., Descamps, N., Souchon, I., & Saint-Eve, A. (2021). Using a mixture
518 design and fraction-based formulation to better understand perceptions of plant-protein-based
519 solutions. *Food Research International*, 110151. <https://doi.org/10.1016/j.foodres.2021.110151>

520 Crevieu-Gabriel, I. (1999). Digestion des protéines végétales chez les monogastriques. Exemple
521 des protéines de pois. *INRA Prod. Anim.*, 12 (2), 147-161.

522 Daher, D., Deracinois, B., Baniel, A., Wattez, E., Dantin, J., Froidevaux, R., Chollet, S., & Flahaut, C.
523 (2020). Principal component analysis from mass spectrometry data combined to a sensory evaluation
524 as a suitable method for assessing bitterness of enzymatic hydrolysates produced from micellar casein
525 proteins. *Foods*, 9(10), 1354. <https://doi.org/10.3390/foods9101354>

526 Davis, J., Sonesson, U., Baumgartner, D. U., & Nemecek, T. (2010). Environmental impact of four
527 meals with different protein sources: Case studies in Spain and Sweden. *Food Research International*,
528 43(7), 1874–1884. <https://doi.org/10.1016/j.foodres.2009.08.017>

529 Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence
530 database search tool. *Proteomics*, 13(1), 22–24. <https://doi.org/10.1002/pmic.201200439>

531 Engel, E., Nicklaus, S., Salles, C., & Le Quéré, J.-L. (2002). Relevance of omission tests to determine
532 flavour-active compounds in food: Application to cheese taste. *Food Quality and Preference*, 13(7),
533 505–513. [https://doi.org/10.1016/S0950-3293\(02\)00136-2](https://doi.org/10.1016/S0950-3293(02)00136-2)

534 Gao, Q., Jiang, H., Tang, F., Cao, H., Wu, X., Qi, F., Sun, J., & Yang, J. (2019). Evaluation of the
535 bitter components of bamboo shoots using a metabolomics approach. *Food & Function*, 10(1), 90–98.
536 <https://doi.org/10.1039/C8FO01820K>

537 Gupta, N. T., Vander Heiden, J. A., Uduman, M., Gadala-Maria, D., Yaari, G., & Kleinstein, S. H.
538 (2015). Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing
539 data: Table 1. *Bioinformatics*, 31(20), 3356–3358. <https://doi.org/10.1093/bioinformatics/btv359>

540 Heng, L., Vincken, J.-P., van Koningsveld, G., Legger, A., Gruppen, H., van Boekel, T., Roozen, J., &
541 Voragen, F. (2006). Bitterness of saponins and their content in dry peas. *Journal of the Science of*
542 *Food and Agriculture*, 86(8), 1225–1231. <https://doi.org/10.1002/jsfa.2473>
543 Huang, W., Shen, Q., Su, X., Ji, M., Liu, X., Chen, Y., Lu, S., Zhuang, H., & Zhang, J. (2016).
544 BitterX: A tool for understanding bitter taste in humans. *Scientific Reports*, 6(1), 23450.
545 <https://doi.org/10.1038/srep23450>
546 Gharibzahedi, S. M. T., & Smith, B. (2021). Effects of high hydrostatic pressure on the quality and
547 functionality of protein isolates, concentrates, and hydrolysates derived from pulse legumes: A review.
548 *Trends in Food Science & Technology*, 107, 466–479. <https://doi.org/10.1016/j.tifs.2020.11.016>
549 Guillot, A., Boulay, M., Chambellon, É., Gitton, C., Monnet, V., & Juillard, V. (2016). Mass
550 spectrometry analysis of the extracellular peptidome of *Lactococcus lactis*: Lines of evidence for the
551 coexistence of extracellular protein hydrolysis and intracellular peptide excretion. *Journal of Proteome*
552 *Research*, 15(9), 3214–3224. <https://doi.org/10.1021/acs.jproteome.6b00424>
553 Iwaniak, A., Minkiewicz, P., Darewicz, M., Sieniawski, K., & Starowicz, P. (2016). BIOPEP database
554 of sensory peptides and amino acids. *Food Research International*, 85, 155–161.
555 <https://doi.org/10.1016/j.foodres.2016.04.031>
556 Jupp, S., Burdett, T., Malone, J., Leroy, C., Pearce, M., & Parkinson, H. (2015). A new ontology
557 lookup service at EMBL-EBI. 2. *SWAT4LS*, 118-119
558 Kato, A., & Nakai, S. (1980). Hydrophobicity determined by a fluorescence probe method and its
559 correlation with surface properties of proteins. *Biochimica et Biophysica Acta (BBA)-Protein*
560 *Structure*, 624(1), 13–20. [https://doi.org/10.1016/0005-2795\(80\)90220-2](https://doi.org/10.1016/0005-2795(80)90220-2)
561 Kim, M.-R., Yukio, K., Kim, K. M., & Lee, C.-H. (2008). Tastes and structures of bitter peptide,
562 Asparagine-Alanine-Leucine-Proline-Glutamate, and its synthetic analogues. *Journal of Agricultural*
563 *and Food Chemistry*, 56(14), 5852–5858. <https://doi.org/10.1021/jf7036664>
564 Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a
565 protein. *Journal of Molecular Biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-](https://doi.org/10.1016/0022-2836(82)90515-0)
566 0

567 Langella, O., Valot, B., Balliau, T., Blein-Nicolas, M., Bonhomme, L., & Zivy, M. (2017).
568 X!TandemPipeline: A Tool to manage sequence redundancy for protein inference and phosphosite
569 identification. *Journal of Proteome Research*, 16(2), 494–503.
570 <https://doi.org/10.1021/acs.jproteome.6b00632>

571 Li, H., & Aluko, R. E. (2010). Identification and inhibitory properties of multifunctional peptides from
572 pea protein hydrolysate. *Journal of Agricultural and Food Chemistry*, 58(21), 11471–11476.
573 <https://doi.org/10.1021/jf102538g>

574 Lioe, H. N., Selamat, J., & Yasuda, M. (2010). Soy sauce and its umami taste: a link from the past to
575 current situation. *Journal of Food Science*, 75(3), R71–R76. [https://doi.org/10.1111/j.1750-](https://doi.org/10.1111/j.1750-3841.2010.01529.x)
576 [3841.2010.01529.x](https://doi.org/10.1111/j.1750-3841.2010.01529.x)

577 Maehashi, K., & Huang, L. (2009). Bitter peptides and bitter taste receptors. *Cellular and Molecular*
578 *Life Sciences*, 66(10), 1661–1671. <https://doi.org/10.1007/s00018-009-8755-9>

579 Maehashi, Kenji, Matsuzaki, M., Yamamoto, Y., & Udaka, S. (1999). Isolation of peptides from an
580 enzymatic hydrolysate of food proteins and characterization of their taste properties. *Biosci Biotechnol*
581 *Biochem*, 63(3), 555–559. <https://doi.org/10.1271/bbb.63.555>

582 Proust, L., Sourabié, A., Pedersen, M., Besançon, I., Haudebourg, E., Monnet, V., & Juillard, V.
583 (2019). Insights into the complexity of yeast extract peptides and their utilization by *Streptococcus*
584 *thermophilus*. *Frontiers in Microbiology*, 10, 906. <https://doi.org/10.3389/fmicb.2019.00906>

585 Roland, W. S. U., Pouvreau, L., Curran, J., van de Velde, F., & de Kok, P. M. T. (2017). Flavor
586 aspects of pulse ingredients. *Cereal Chemistry Journal*, 94(1), 58–65.
587 <https://doi.org/10.1094/CCHEM-06-16-0161-FI>

588 Salger, M., Stark, T. D., & Hofmann, T. (2019). Taste modulating peptides from overfermented Cocoa
589 Beans. *Journal of Agricultural and Food Chemistry*, 67(15), 4311–4320.
590 <https://doi.org/10.1021/acs.jafc.9b00905>

591 Sarmadi, B. H., & Ismail, A. (2010). Antioxidative peptides from food proteins: A review. *Peptides*,
592 31(10), 1949–1956. <https://doi.org/10.1016/j.peptides.2010.06.020>

593 Sirtori, E., Isak, I., Resta, D., Boschin, G., & Arnoldi, A. (2012). Mechanical and thermal processing
594 effects on protein integrity and peptide fingerprint of pea protein isolate. *Food Chemistry*, 134(1),
595 113–121. <https://doi.org/10.1016/j.foodchem.2012.02.073>

596 Su, G., Cui, C., Zheng, L., Yang, B., Ren, J., & Zhao, M. (2012). Isolation and identification of two
597 novel umami and umami-enhancing peptides from peanut hydrolysate by consecutive chromatography
598 and MALDI-TOF/TOF MS. *Food Chemistry*, 135(2), 479–485.
599 <https://doi.org/10.1016/j.foodchem.2012.04.130>

600 Sun, J., He, H., & Xie, B. J. (2004). Novel antioxidant peptides from fermented mushroom
601 *Ganoderma lucidum*. *Journal of Agricultural and Food Chemistry*, 52(21), 6646–6652.
602 <https://doi.org/10.1021/jf0495136>

603 Temussi, P. A. (2012). The good taste of peptides: peptides taste. *Journal of Peptide Science*, 18(2),
604 73–82. <https://doi.org/10.1002/psc.1428>

605 Toelstede, S., & Hofmann, T. (2008). Sensomics mapping and identification of the key bitter
606 metabolites in Gouda cheese. *Journal of Agricultural and Food Chemistry*, 56(8), 2795–2804.
607 <https://doi.org/10.1021/jf7036533>

608 Valot, B., Langella, O., Nano, E., & Zivy, M. (2011). MassChroQ: A versatile tool for mass
609 spectrometry quantification. *Proteomics*, 11, 3572–3577. <https://doi.org/10.1002/pmic.201100120>

610 Wang, W., Zhou, X., & Liu, Y. (2020). Characterization and evaluation of umami taste: A review.
611 *Trends in Analytical Chemistry*, 127, 115876. <https://doi.org/10.1016/j.trac.2020.115876>

612 Zimmerman, J. M., Eliezer, N., & Simha, R. (1968). The characterization of amino acid sequences in
613 proteins by statistical methods. *Journal of Theoretical Biology*, 21(2), 170–201.
614 [https://doi.org/10.1016/0022-5193\(68\)90069-6](https://doi.org/10.1016/0022-5193(68)90069-6)

615

616 **Table 1:** Overall characteristics of the six fractions and the two raw solutions

	Number of peptides identified	Sum of peptide area	Dry matter content (% w/w)	Protein content (%)	Ash content (%)	Conductivity (mS/cm) at 20°C	pH at 20°C	Surface hydrophobicity index
100Pa	2586	1.31E+11	0.20	0.04	0.07	1.44	8.4	363
100Pb	1756	4.92E+10	0.20	0.04	0.04	1.16	9.3	298
100Ra	2376	7.05E+10	1.70	1.41	0.15	1.08	7.5	933
100Rb	1551	2.98E+10	1.70	1.48	0.12	0.88	7.5	1269
50Ia-50W	1565	3.72E+10	6.00	4.91	0.18	1.06	7.5	2083
50Ib-50W	809	9.26E+09	6.00	5.10	0.19	0.84	7.5	2172
Refa	2235	7.13E+10	94.00	79.05	4.14	1.09	7.5	2961
Refb	1488	2.26E+10	93.70	80.68	3.84	1.01	7.5	3504

617

618

619 Captions to Figures

620

621 **Figure 1:** General workflow of the different steps of peptidomic analysis: the preparation and
622 measurement processes (in orange) to the bioinformatic analyses (in blue), the preprocessing of the
623 data (in green), and the calculations for the recombined products (in yellow).

624

625 **Figure 2:** Categorization of the 3,005 unique peptides identified via UHPLC-MS/MS based on protein
626 origin (threshold for peptide number: 24).

627

628 **Figure 3:** A) Distribution of the overall 3,561 peptides on the 28 solutions (normalized distribution,
629 kernel density): A1 = Relative aromatic nature; A2 = Relative acidic nature; A3 = Aliphatic index; A4
630 = Relative basic nature; A5 = Polarity; A6 = Bulk; A7 = GRAVY index; A8 = Length; A9 = Charge.

631 B) Comparison of distributions of the overall 3,561 peptides (normalized distribution, kernel density):
632 B1 = Charge with 50Ib-50W in blue; 100Rb in green and 100Pb in red; B2 = Charge with 50Ia-50W
633 in blue; 100Ra in green and 100Pa in red; B3 = Polarity with Refa in yellow and Refb in orange.

634 C) Comparison of distributions of the overall peptides (3,561 peptides in blue) and for the peptides
635 correlated to bitterness (275 peptides in red) on the 28 solutions (normalized distribution, kernel
636 density): C1 = Polarity; C2 = Aliphatic index; C3 = GRAVY index.

637

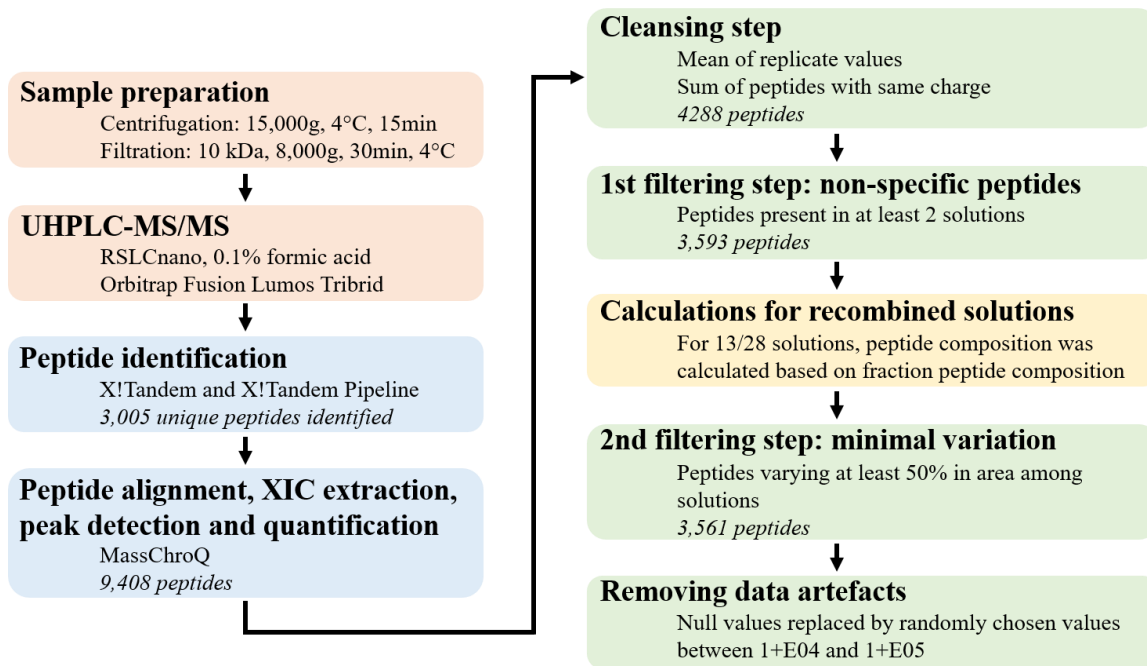
638 **Figure 4:** Depiction of the intersections in peptide sets among the six fractions and the two raw
639 solutions (UpSet plot). The blue horizontal bars show the number of peptides in each fraction/solution.
640 The black dots and lines show the combinations of peptides that make up each cluster or subset of the
641 fractions/solutions. The vertical histogram shows the number of peptides in each subset.

642

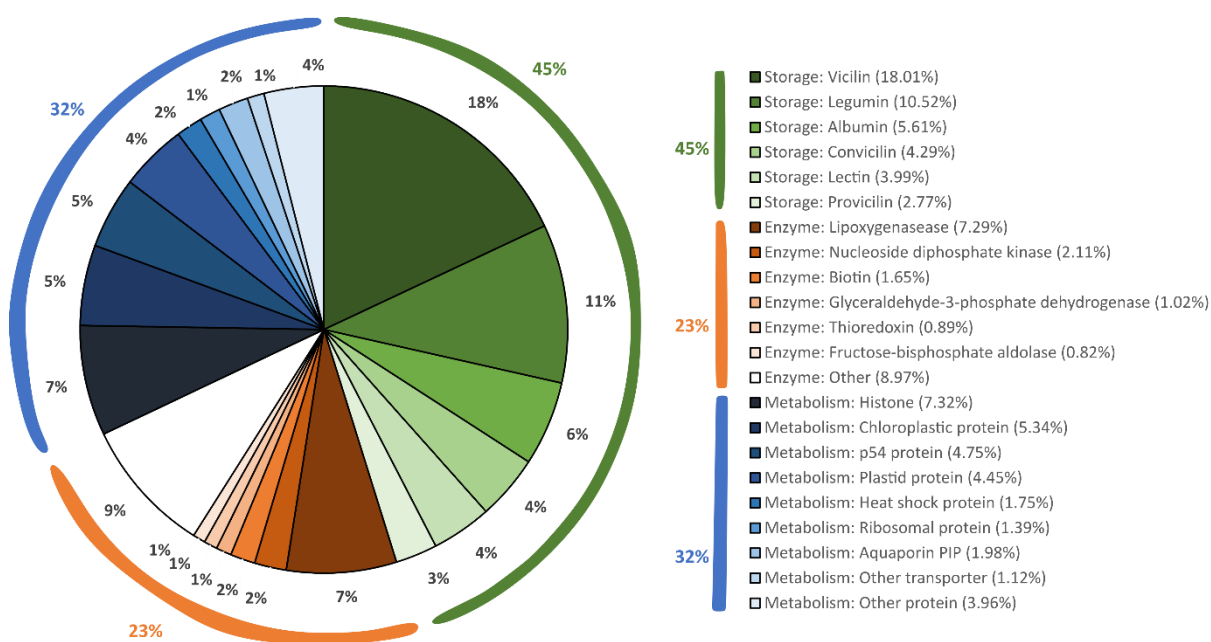
643 **Figure 5:** Results of the principal component analysis (PCA, wide method) examining the peptide
644 profiles for the 28 solutions, which were determined using a fraction-based formulation strategy and
645 the peptides that had been identified (based on 3,561 peptides). On the left is a loading plot showing
646 the correlational relationships between the PCA axes 1 and 2 and the peptide areas. On the right is a
647 PCA plot with the same two axes that shows the relative similarity in the solutions' peptide profiles.
648 The green circles are the recombined solutions created from batch b. The blue triangles are the
649 recombined solutions created from batch a. The orange squares are the recombined solutions created
650 from batch a and batch b. The dark star is the water solution. The solid symbols represent the
651 measured values, and the empty symbols represent the calculated values.

652

653 **Figure 1:** General workflow of the different steps of peptidomic analysis: the preparation and
 654 measurement processes (in orange) to the bioinformatic analyses (in blue), the preprocessing of the
 655 data (in green), and the calculations for the recombined products (in yellow).
 656



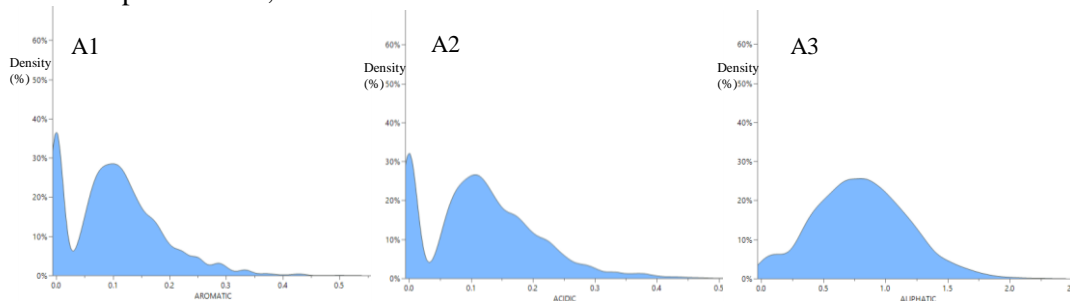
657
 658 **Figure 2:** Categorization of the 3,005 unique peptides identified via UHPLC-MS/MS based on protein
 659 origin (threshold for peptide number: 24).
 660



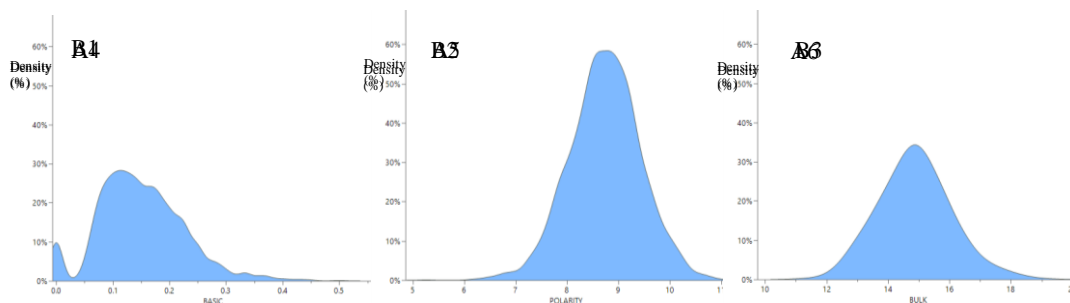
661
 26

663 **Figure 3** : A) Distribution of the overall 3,561 peptides on the 28 solutions (normalized distribution, kernel
 664 density): A1 = Relative aromatic nature; A2 = Relative acidic nature; A3 = Aliphatic index; A4 = Relative
 665 basic nature; A5 = Polarity; A6 = Bulk; A7 = GRAVY index; A8 = Length; A9 = Charge.
 666 B) Comparison of distributions of the overall 3,561 peptides (normalized distribution, kernel density): B1 =
 667 Charge with 50Ib-50W in blue; 100Rb in green and 100Pb in red; B2 = Charge with 50Ia-50W in blue; 100Ra
 668 in green and 100Pa in red; B3 = Polarity with Refa in yellow and Refb in orange.
 669 C) Comparison of distributions of the overall peptides (3,561 peptides in blue) and for the peptides correlated
 670 to bitterness (275 peptides in red) on the 28 solutions (normalized distribution, kernel density): C1 = Polarity;
 671 C2 = Aliphatic index; C3 = GRAVY index.

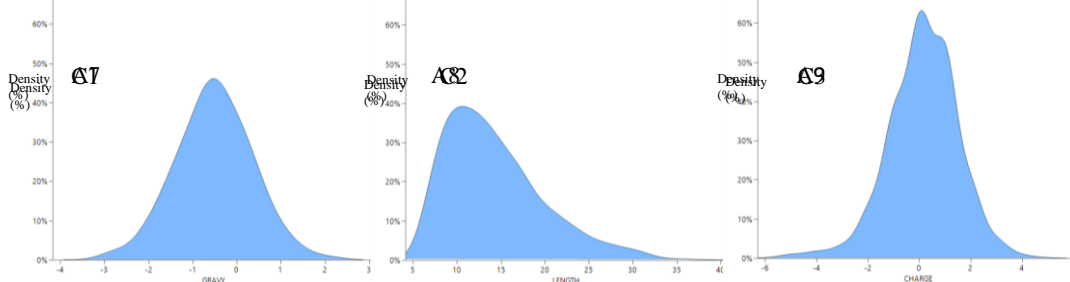
672



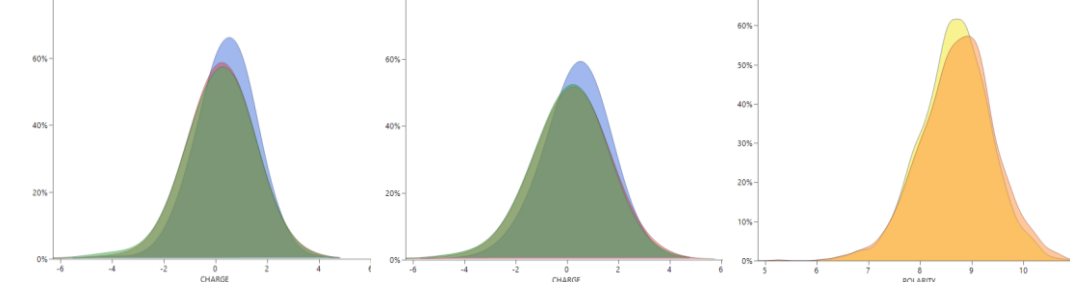
673



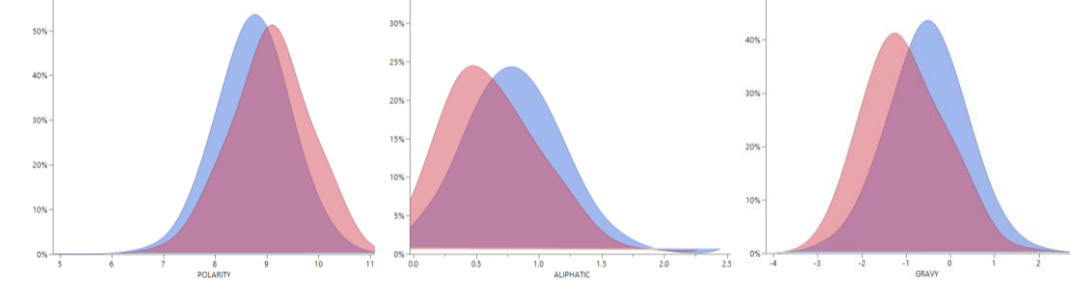
674



675

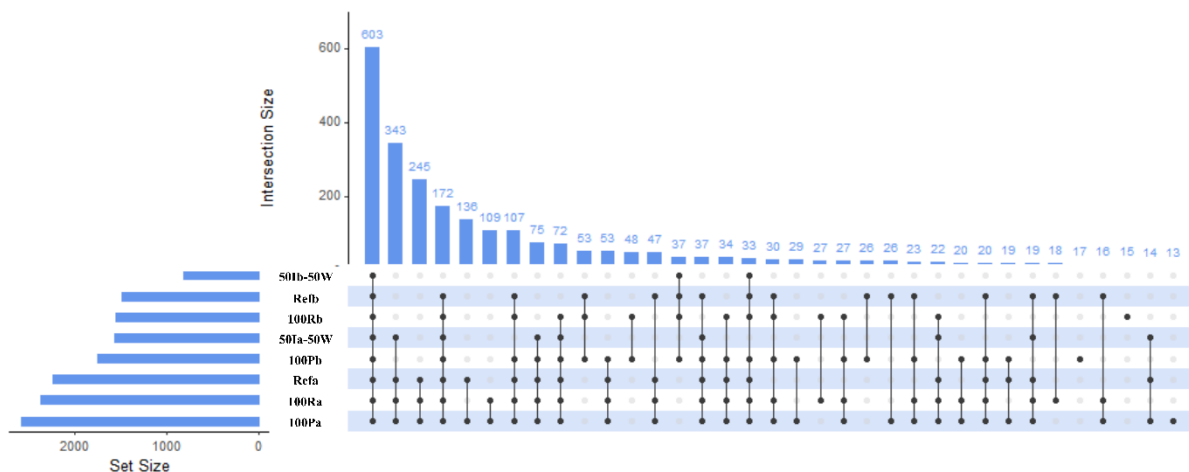


676



677 **Figure 4:** Depiction of the intersections in peptide sets among the six fractions and the two raw solutions
 678 (UpSet plot). The blue horizontal bars show the number of peptides in each fraction/solution. The black dots
 679 and lines show the combinations of peptides that make up each cluster or subset of the fractions/solutions. The
 680 vertical histogram shows the number of peptides in each subset.

681
 682



683
 684

685

686

687

688

689

690

691

692

693

694

695

696

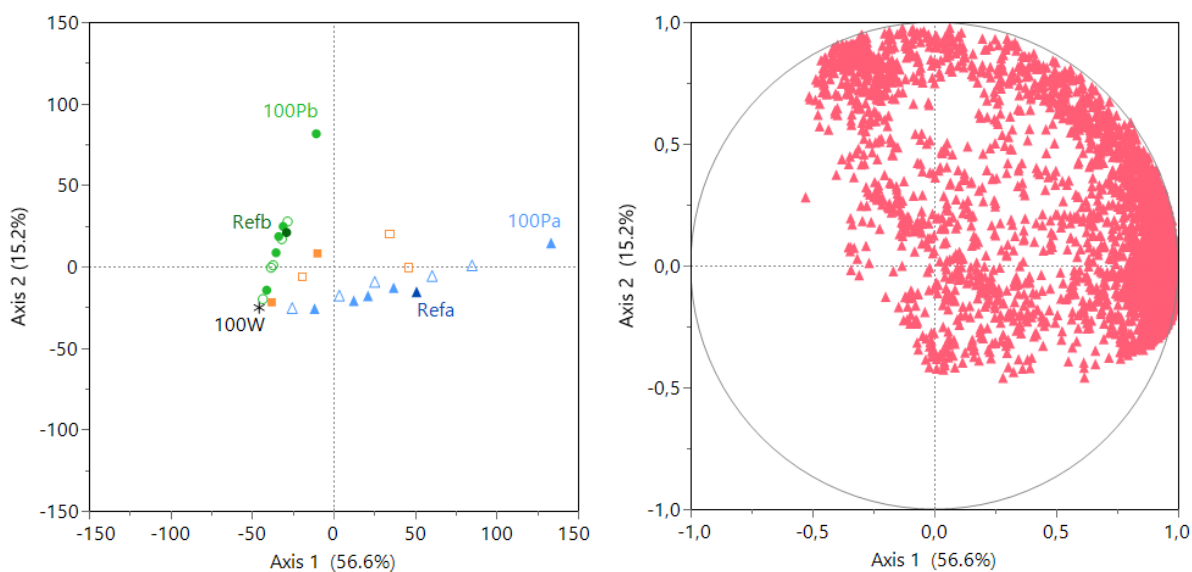
697

698

699

700

701 **Figure 5:** Results of the principal component analysis (PCA, wide method) examining the peptide profiles for
702 the 28 solutions, which were determined using a fraction-based formulation strategy and the peptides that had
703 been identified (based on 3,561 peptides). On the left is a loading plot showing the correlational relationships
704 between the PCA axes 1 and 2 and the peptide areas. On the right is a PCA plot with the same two axes that
705 shows the relative similarity in the solutions' peptide profiles. The green circles are the recombined solutions
706 created from batch b. The blue triangles are the recombined solutions created from batch a. The orange
707 squares are the recombined solutions created from batch a and batch b. The dark star is the water solution. The
708 solid symbols represent the measured values, and the empty symbols represent the calculated values.
709



710

711

712

713 **Captions to Supplementary Tables**

714

715 **Supplementary Table 1:** Composition of the 28 solutions used in this study, which were created by mixing
716 permeates a and b, retentates a and b, and pellets a and b. For the coding: Refa (respectively Refb) correspond
717 to the solutions of pea protein isolates a (resp. b) at 4% (w/w); “Pa” (resp. “Pb”) mean permeate from Refa
718 (resp. from Refb), “Ia” (resp. “Ib”) mean Pellet from Refa (resp. from Refb), “Ra” (resp. “Rb”) mean
719 Retentate from Refa (resp. from Refb) and “W” mean water. For example, “X Pb-Y Ra” mean “Recombined
720 product constituted of X% of permeate from Refb and Y% of retentate from Refa”.

721

722 **Supplementary Table 2:** Peptides identified in the pea protein solutions that had sequences homologous to
723 those of previously described antioxidant peptides (BIOPEP database) (Iwaniak et al., 2016).

724

725

726 **Supplementary Table 1:**

Solution ID	Permeate a (%)	Permeate b (%)	Retentate a (%)	Retentate b (%)	Pellet a (%)	Pellet b (%)	Water (%)
100W	0	0	0	0	0	0	100
25Ib-75W	0	0	0	0	0	25	75
50Ib-50W	0	0	0	0	0	50	50
30Ia-70W	0	0	0	0	30	0	70
50Ia-50W	0	0	0	0	50	0	50
40Rb-30Ib-30W	0	0	0	40	0	30	30
40Rb-30Ia-30W	0	0	0	40	30	0	30
50Rb-50W	0	0	0	50	0	0	50
100Rb	0	0	0	100	0	0	0
40Ra-30Ib-30W	0	0	40	0	0	30	30
50Ra-25Ia-25W	0	0	50	0	25	0	25
60Ra-40W	0	0	60	0	0	0	40
100Ra	0	0	100	0	0	0	0
40Pb-60W	0	40	0	0	0	0	60
Refb	0	40	0	36	0	24	0
50Pb-50W	0	50	0	0	0	0	50
50Pb-25Ib-25W	0	50	0	0	0	25	25
50Pb-50Rb	0	50	0	50	0	0	0
70Pb-30Ra	0	70	30	0	0	0	0
100Pb	0	100	0	0	0	0	0
25Pa-25Ra-13Ia-38W	25	0	25	0	12.5	0	37.5
Refa	38	0	34	0	28	0	0
40Pa-60W	40	0	0	0	0	0	60
40Pa-60Rb	40	0	0	60	0	0	0
50Pa-25Ib-25W	50	0	0	0	0	25	25
50Pa-Ia25-W25	50	0	0	0	25	0	25
50Pa-50Ra	50	0	50	0	0	0	0
100Pa	100	0	0	0	0	0	0

727

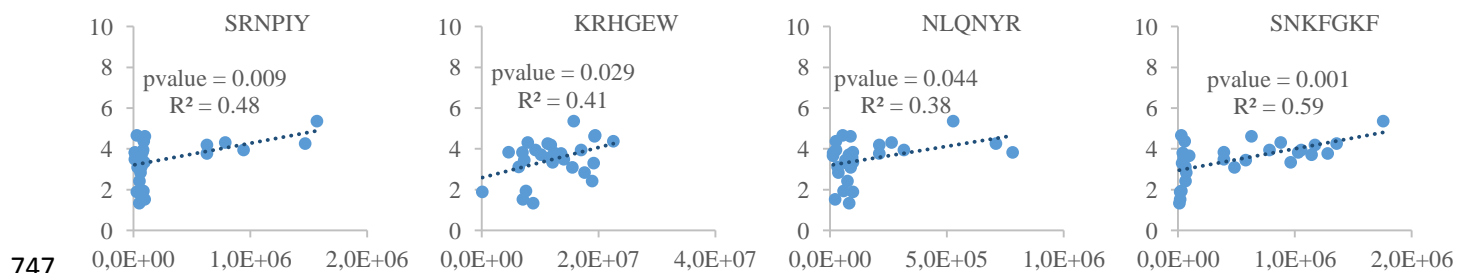
728 **Supplementary Table 2:**

Antioxidative sequence	Protein	Peptide identified in this study
ADGF	Lectin	SYNVADGFTFF
		VINAPNSYNVADGFT
		VINAPNSYNVADGFTF
		VINAPNSYNVADGFTFF
ADVFNPR	Legumin L1 beta chain	HEDLAGSSQADVFNPRAGRIT
		HEDLAGSSQADVFNPRAGRITSVN
		HEDLAGSSQADVFNPRAGRITSVNSLT
		HEDLAGSSQADVFNPRAGRITSVNSLTL
		HEDLAGSSQADVFNPRAGRITSVNSLTLPLVK
		HEDLAGSSQADVFNPRAGRITSVNSLTLPLVKL
		LKLHEDLAGSSQADVFNPRAGRITSVN
		LKLHEDLAGSSQADVFNPRAGRITSVNSLT
ELLI	Histone H3.2	STELLIR
FVPH	PsRT17-1	VFVPHIRTLGD
		VFVPHIRTLGDA
FVPH and SAEHGSLH	Legumin A2	SAEHGSLHKNAM(MOD:00719)FVPH
		SAEHGSLHKNAM(MOD:00719)FVPHY
HLHP	Sucrose transport protein	QLSGAFKELKRPM(MOD:00719)W
KFPE	PIP1-2	M(MOD:00719,MOD:00408)EAKEEDVSLGANKFPERQPIG
		M(MOD:00719,MOD:00408)EAKEQDVSLGANKFPERQPLG
KFPE	PIP-type 7a	M(MOD:00408)EAKEQDVSLGANKFPERQPLG
LPILR	Legumin (Minor small)	LPILRN
		LPILRNL
		SGAGRISTVNSLTLPILR
		SGAGRISTVNSLTLPILRN
		SGAGRISTVNSLTLPILRNL
SGAF	Malate dehydrogenase	Q(MOD:01160)RIARISAPHLHPSN
YLKT	Seed linoleate 9S-lipoxygenase-3	VKSPQKAYLKTITP
		VKSPQKAYLKTITPKFQT
		YLKTITP
YVGD	Actin-3	AYVGDQAQSKRGILT

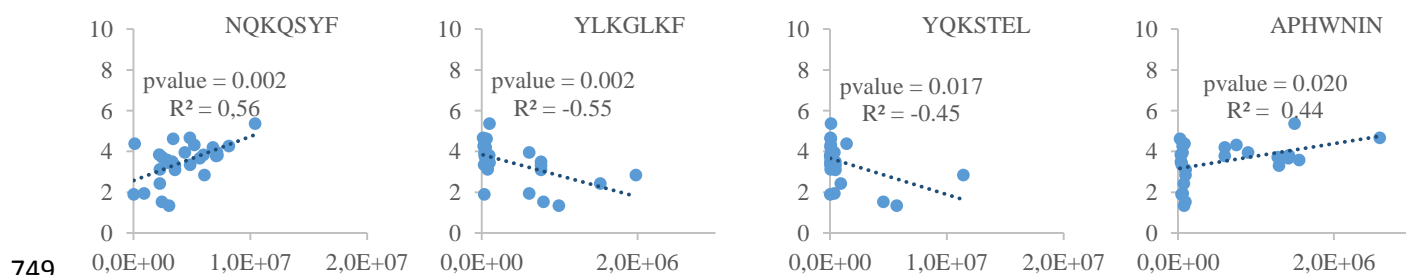
729

744 **Supplementary Figure 3: Peptides (size < 8 residues) correlated with perceived bitterness (score out of 10);**
 745 the R^2 values and p-values from the Pearson's correlational analysis are indicated

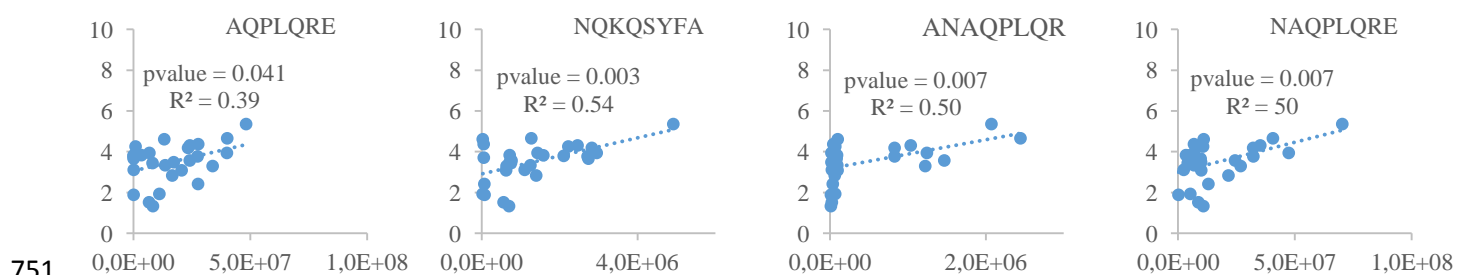
746



748



750



752

753

754

