Elsevier Editorial System(tm) for Food Chemistry or its open access mirror
Manuscript Draft

Manuscript Number: FOODCHEM-D-20-04981

Title: Comprehensive Identification of Native Medium-Sized and Short Bioactive Peptides in Sea Bass Muscle

Article Type: Research Article (max 7,500 words)

Keywords: peptidomics; native peptides; short peptides; bioactive peptides; fish

Corresponding Author: Dr. Susy Piovesana, Ph.D

Corresponding Author's Institution: Sapienza Università di Roma

First Author: Andrea Cerrato

Order of Authors: Andrea Cerrato; Sara Elsa Aita; Chiara Cavaliere, Ph.D.; Aldo Laganà; Carmela Maria Montone; Susy Piovesana; Riccardo Zenezini Chiozzi; Anna Laura Capriotti

Abstract: Native peptides from sea bass muscle were analyzed by two different approaches: medium-sized peptides by peptidomics analysis, whereas short peptides by suspect screening analysis employing an inclusion list of exact m/z values of all possible amino acid combinations (from 2 up to 4). The method was also expanded to common post-translational modifications potentially interesting in food analysis, as well as non-proteolytic aminoacyl derivatives, which are already well-known taste-active building blocks in pseudo-peptides. The medium-sized peptides were identified by de novo and combination of de novo and spectra matching to protein sequence database, with up to 4077 peptides (2725 modified) identified by database search and 2665 peptides (223 modified) identified by de novo only; 102 short peptide sequences were identified (with 12 modified ones), and most of them had multiple reported bioactivities.
The method can be extended to any peptide mixture, either endogenous or by protein hydrolysis, from other food matrices.

Dear Editor,

I am submitting the manuscript entitled "Comprehensive Identification of Native Medium-Sized and Short Bioactive Peptides in Sea Bass Muscle" with the following co-authors: Andrea Cerrato, Sara Elsa Aita, Chiara Cavaliere, Aldo Laganà, Carmela Maria Montone, Riccardo Zenezini Chiozzi, Anna Laura Capriotti.

The manuscript describes an analytical procedure for the comprehensive characterization of peptides from complex extracts. The method is particularly suited to discover new peptides in food and related matrices for identifications of potential bioactive peptides. By this approach the complete peptidome is addressed, which does not only comprise the peptides which can be usually investigated by peptidomics, but also very short peptide sequences (dipeptides, tripeptides, tetrapeptides). Such short peptides are particularly interesting, as most of them are known bioactive peptides, but poorly characterized due to analytical workflows which are not specifically designed for their special needs. Therefore, in this work we describe two parallel analytical procedures, one based on peptidomics for the characterization of middle-sized peptides, one specific for short peptides. In particular, the latter procedure comprised a clean-up on graphitized carbon black to enrich short peptides, a suspect screening MS acquisition strategy with inclusion list to tackle low abundance and poor ionization of short peptides, a data analysis based on Compound Discoverer software to screen the acquired MS data for the short peptides potentially present in the sample and aid and speed up peptide identification from MS spectra. By this approach, more 102 short peptides were indentified from sea bass extract of native peptides present in fish fillet, along with more than 4000 middle-sized peptides obtained by a peptidomic strategy. Finally, the study also aimed at characterizing peptides with modifications, especially the ones which have been recognized or suggested as interesting in food, for bioactivity or taste. Currently such peptides are very poorly studied and the present method can aid in this, filling a gap and providing new information useful to develop nutraceuticals with better taste and bioactivity.

The work used sea bass fillet as sample, but the analytical procedure can be easily modified and adapted to any peptide extract, being it of endogenous peptides or obtained by digestion of precursor proteins.

With kindest regards,

June 10th, 2020

Susy Piovesana
Department of Chemistry
Università di Roma "La Sapienza"
e-mail: susy.piovesana@uniroma1.it

*Highlights (for review)

Highlights

- Native peptides and their modifications were identified in sea bass fillet extract
- Two parallel workflows to meet the requirements of middle-sized and short peptides
- 102 short peptides (dipeptides, tripeptides) were confidently identified
- More than 4000 middle-sized peptides were identified

*Manuscript
Click here to download Manuscript: Manuscript.docx

1 **Comprehensive Identification of Native Medium-Sized and Short Bioactive Peptides in**
2 **Sea Bass Muscle**
3
4 Andrea Cerrato[1], Sara Elsa Aita[1], Chiara Cavaliere[1], Aldo Laganà[1,2], Carmela Maria
5 Montone[1], Susy Piovesana*[1], Riccardo Zenezini Chiozzi[3,4], Anna Laura Capriotti[1]
6
7 1. Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185
8 Rome, Italy
9 2. CNR NANOTEC, Campus Ecotekne, University of Salento, Via Monteroni, 73100 Lecce,
10 Italy
11 3. Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular
12 Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8,
13 3584 CH Utrecht, The Netherlands.
14 4. Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.
15
16 ***e-mail addresses:***
17 *andrea.cerrato@uniroma1.it; saraelsa.aita@uniroma1.it; chiara.cavaliere@uniroma1.it;*
18 *aldo.lagana@uniroma1.it; carmelamaria.montone@uniroma1.it;*
19 *susy.piovesana@uniroma1.it; r.zenezinichiozzi@uu.nl; annalaura.capriotti@uniroma1.it*
20
21 ***Corresponding author**: Susy Piovesana
22 Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome,
23 Italy
24 E-mail: susy.piovesana@uniroma1.it
25 Telephone: (+39) 06 4991 3062

Abstract

Native peptides from sea bass muscle were analyzed by two different approaches: medium-sized peptides by peptidomics analysis, whereas short peptides by suspect screening analysis employing an inclusion list of exact *m*/*z* values of all possible amino acid combinations (from 2 up to 4). The method was also expanded to common post-translational modifications potentially interesting in food analysis, as well as non-proteolytic aminoacyl derivatives, which are already well-known taste-active building blocks in pseudo-peptides. The medium-sized peptides were identified by *de novo* and combination of *de novo* and spectra matching to protein sequence database, with up to 4077 peptides (2725 modified) identified by database search and 2665 peptides (223 modified) identified by *de novo* only; 102 short peptide sequences were identified (with 12 modified ones), and most of them had multiple reported bioactivities.

The method can be extended to any peptide mixture, either endogenous or by protein hydrolysis, from other food matrices.

Keywords

peptidomics; native peptides; short peptides; bioactive peptides; fish

1. Introduction

The valorisation of food health benefits and the identification of nutraceutical compounds, suitable for functional food production, are one of the main goals to support a healthy diet and lifestyle in disease prevention. Along with the studies for the characterization of differences between farmed and wild animals (Zenezini Chiozzi et al., 2018), bioactive peptides are particularly interesting in this context, because they go beyond the dietary function by displaying a variety of bioactivities with health promoting effects. Bioactive peptides can be endogenous components in ordinary food (the peptides are integral part of the food composition or they are produced by processing, maturation, storage, etc.), but they also can be produced by digestion in the body after consuming food (Capriotti, Caruso, et al., 2015) or prepared artificially by proteases and fermentation processes, as in the case of fermented food, biomasses (Montone et al., 2018b) and for the waste revalorization (Montone et al., 2018a; Montone, Zenezini Chiozzi, et al., 2019).

Fish has been recognized and extensively studied for discovery and identification of new bioactive peptides. Along with the nutritional purposes, fish represents a sources of unique peptides with a broad spectrum of biological activities, as demonstrated by studies on the pharmacological effects of fish-derived peptides, which were proved to exhibit antihypertensive (Yathisha et al., 2019), immunomodulatory, antioxidant, antitumor and antimicrobial (Capriotti, Cavaliere, et al., 2015) activities (Valero et al., 2020). Not only fish fillets, but also by-products were thoroughly investigated and proteins there recovered and hydrolysed with different proteases to search for possible valuable compounds with bioactive properties: for example, peptides with antioxidant and cytotoxic activity to cancer cells were obtained from Flathead (*Platycephalus fuscus*) head, backbone and frames (Nurdiani et al., 2017).

67 Though the investigation of bioactive peptides in fish is mainly achieved on protein
68 hydrolysates, endogenous peptides are expected to be present as well. Endogenous peptides
69 in food are mainly characterized in other matrices, such as meat (aged duck meat (D. Liu et
70 al., 2019), fermented meat source, chicken breast and ham (R. Liu et al., 2016)), milk
71 (Capriotti et al., 2016; Nongonierma & FitzGerald, 2016; Zenezini Chiozzi et al., 2016),
72 vegetables (Piovesana et al., 2018), and have already been demonstrated to be potentially
73 valuable bioactive compounds. As most bioactive sequences are five or less amino acid long,
74 the specific issues of their isolation and identification was also in some instances considered,
75 especially to investigate milk-derived peptides (Montone, Capriotti, et al., 2019; O'Keeffe &
76 Fitzgerald, 2015), but they are emerging as fundamental in other matrices as well, in
77 particular in dry-cured ham, where dipeptides and tripeptides are not only abundant, but also
78 display multiple bioactivities and important contribution to taste. As such, ham is one of the
79 most investigated matrices for naturally occurring bioactive peptides (Mora et al., 2017),
80 which are produced due to the intense proteolysis of muscle proteins by the action of
81 endogenous enzymes and also microbial peptidases in the case of dry-fermented meats. In
82 ham, most discovered bioactive peptides are short sequences, between 2 to 10 amino acid
83 long (Mora et al., 2017). For instance, in Spanish dry-cured ham endogenous peptides were
84 investigated by size-exclusion chromatography and peptidomics technologies, indicating
85 them to be relatively small (below 20 amino acids and most of them being 5 amino acid long)
86 and with antioxidant activity (Mora et al., 2014). Such peptides were also demonstrated to
87 have a dipeptidyl peptidase IV (DPP-IV) inhibitor activity, potentially beneficial in type 2
88 diabetes (Gallego et al., 2014) and activity against the pathogenic microorganisms *Listeria*
89 *monocytogenes* (Castellano et al., 2016). Similarly, Xuanwei ham endogenous peptides were
90 investigated by a combination of size exclusion chromatography, anion exchange
91 chromatography and reversed phase (RP) chromatography, and showed antioxidant activity,

92 mainly attributed to a tetrapeptide (Xing et al., 2016), while Iberian ham was recently
93 demonstrated to be a potential source of α-glucosidase-inhibitory peptides, suitable for type 2
94 diabetes mellitus management (Mora et al., 2020).
95 In fish, the investigation of endogenous peptides is far less common than in meat and mostly
96 carried out to investigate post-mortem changes in organoleptic properties of fish, though
97 there are some exceptions. The activity of endogenous proteases, especially calpains and
98 cathepsins, has been associated with hydrolysis of myofibrillar proteins, resulting not only in
99 softening of fish texture, but also in the production of bioactive peptides (Ahmed et al.,
100 2015). More in detail, the proteases involved in the process are mainly aspecific or with little
101 specificity, and include the endogenous muscle proteases (cytosolic calpains, both m- and μ-
102 calpains, lysosomal cathepsins B, H, and L as well as the aspartic cathepsin D) and
103 connective tissue proteases (elastase and collagenase). The autolytic process is also
104 influenced by physicochemical conditions including low temperature, reduced pH and
105 relatively increased muscle ionic strength. The products of this endogenous proteolytic
106 activity include polypeptide fragments, oligopeptides, but also short peptides and amino
107 acids. The latter can be produced by cathepsin B, for which a strong dipeptidyl
108 carboxypeptidase (peptidyl-dipeptidase) activity was reported, which results is production of
109 dipeptides cleaved from the C-terminal of proteins and peptides. Such peptides, released in
110 fish during storage, can then be consumed as food components.
111 There have been studies investigating the bioactivity of fish endogenous peptides, which
112 indicated interesting activities of these extracts, including antioxidant activity for Pacific hake
113 (*M. productus*) muscle, antihypertensive activity (angiotensin-converting enzyme (ACE)
114 inhibitor activity) for sardinelle (*S. aurita*) by-products, Australian salmon, barracouta, and
115 silver warehou, satiety enhancer activity of smooth hound (*M. mustelus*), α-amylase

116 inhibition and antidiabetic activity of Australian salmon, barracouta, and silver warehou
117 (Ahmed et al., 2015).
118 Given the above, the aim of this work was to develop a comprehensive methodology for
119 investigation of peptides, both small and medium-sized, in European sea bass (*Dicentrarchus*
120 *labrax*) fillet samples. Native endogenous peptides were targeted and two fractions obtained,
121 one for medium-sized peptides and one specifically addressing the issue of isolation and
122 purification of short peptide sequences. Similarly, analysis and identification were considered
123 as well, using a peptidomic strategy for identification of medium-sized peptides which
124 included *de novo* identification and combination of it with spectra matching to protein
125 sequence database, and a suspect screening investigation specific for short peptides. Post-
126 translational modifications (PTMs) were also included in the investigation, as they are rarely
127 considered but potentially interesting for a better characterization of the bioactivity and taste
128 of peptide samples.
129
130 **2. Experimental section**
131 **2.1. Materials**
132 All chemicals, reagents and organic solvents of the highest grade available were purchased
133 from Sigma-Aldrich (St. Louis, MO, USA) unless otherwise stated. Trifluoroacetic acid
134 (TFA) was supplied by Romil Ltd (Cambridge). Ultrapure water was prepared by arium 611
135 VF system from Sartorius (Göttingen, Germany). Mass grade solvents used for medium-sized
136 peptides were purchased from VWR International (Milan, Italy). Optima® LC-MS grade
137 water and acetonitrile (ACN), used for short peptide analysis, were purchased from Thermo
138 Fisher Scientific (Waltham, Massachusetts, USA). Bond elut C18 EWP cartridges (1 g) were
139 purchased from Agilent (Santa Clara, USA). Graphitized carbon black (GCB) cartridges

140 packed with 500 mg of Carbograph 4 (130 m^2/g surface area, 20/400-120/200 mesh size)
141 were purchased from Lara S.r.l. (Formello, RM, Italy).
142

143 **2.2. Fish Native Peptide Extraction**

144 Sea bass (*Dicentrarchus labrax*) samples were purchased from a local fish shop and skin,
145 bones and entrails were eliminated to prepare fillets, then aliquoted and store at -20 °C.
146 Twenty g of muscle sample was ground to fine powder with mortar, pestle and the aid of
147 liquid nitrogen. Then, 10 g of the powder was extracted with 20 mL of cold buffer consisting
148 of 100 mmol L^{-1} Tris (hydroxymethyl) aminomethane (Tris-HCl, pH 8). The sample was
149 vortexed for 50 min, then centrifuged at 9400 × g at 4 °C for 30 min to sediment debris, and
150 the supernatant transferred into a new tube. The proteins were precipitated by adding three
151 volumes of ACN and placing the samples at 4 °C for 3 hours. The samples were centrifuged
152 as previously described to recover the supernatant containing the native peptides. The
153 supernatant was transferred to a new tube and ACN was removed under nitrogen flow. The
154 sample was acidified with TFA. Three experimental replicates were performed. Each extract
155 was dived in two aliquots, which were enriched for medium-sized and short peptides,
156 respectively.
157

158 **2.3. Native Peptide Purification**

159 **2.3.1. Purification of Medium-Sized Peptides**

160 Solid-phase extraction (SPE) on C18 was used to concentrate medium-sized peptides from
161 fish extracts. Bond elut C18 EWP cartridges were preliminary washed with 15 mL of ACN
162 and conditioned with 15 mL of 0.1% TFA; then the fish extract was loaded onto the cartridge
163 and washed with 15 mL of 0.1% TFA. Peptides were eluted from the SPE column with 15
164 mL ACN/H_2O (50:50, *v/v*) containing 0.1% TFA, and were dried in a SpeedVac SC250

165 Express (Thermo Savant, Holbrook, NY, USA). The residue was reconstituted in 1 mL of
166 0.1% formic acid in H_2O, filtered through a nylon syringe filters (Pall Acrodisc nylon, 13 mm
167 diameter, 0.45 µm pore diameter, VWR International). As reported above, three experimental
168 replicates were performed.
169

170 **2.3.2. Purification of Short Peptides**

171 For SPE of short peptides, the fish extract with native peptides was loaded on a 500 mg GCB
172 cartridge and short peptides isolated as previously described (Piovesana, Capriotti, et al.,
173 2019). The adsorbent was previously washed with 5 mL of CH_2Cl_2/MeOH, 80:20 (*v*/*v*) with
174 20 mmol L^{-1} TFA and 5 mL of MeOH with 20 mmol L^{-1} TFA. Then, the material was
175 activated by flushing 10 mL of 0.1 mol L^{-1} HCl and finally conditioned with 10 mL of 20
176 mmol L^{-1} TFA. After loading of the fish sample, the cartridge was washed with 2 mL of 20
177 mmol L^{-1} TFA and 0.5 mL of MeOH (used to remove the cartridge dead volume). Finally,
178 analytes were eluted in back-flushing mode by 10 mL of CH_2Cl_2/MeOH, 80:20 (*v*/*v*) with 20
179 mmol L^{-1} TFA. The eluate was evaporated at room temperature in the Speed-Vac and the
180 residue reconstituted in 100 µL H_2O. As reported above, three experimental replicates were
181 performed.
182

183 **2.4. Chromatography-Mass Spectrometry Analysis of Peptide Samples**

184 **2.4.1. Analysis of Medium-Sized Peptides by nano High Performance Liquid**
185 **Chromatography-MS/MS**

186 Medium-sized peptides were analyzed by nanoHPLC on an Ultimate 3000 (Thermo
187 Scientific, Bremen, Germany) coupled to an Orbitrap Elite mass spectrometer (Thermo
188 Scientific). Twenty µL were injected and preconcentrated on a µ-precolumn (Thermo Fischer
189 Scientific, 300 µm i.d. × 5 mm Acclaim PepMap 100 C18, 5 µm particle size, 100 Å pore

190 size) at 10 μL min^{-1} flow rate of a premixed mobile phase H_2O/ACN 98:2 (*v*/*v*) containing
191 0.1% (*v*/*v*) TFA. Then, samples were separated on an EASY-Spray column (Thermo Fischer
192 Scientific, 15 cm × 75 μm i.d. PepMap C18, 3 μm particles, 100 Å pore size) operated at 250
193 nL min^{-1} and at 20 °C. A 55 min-long gradient was employed with H_2O and ACN as mobile
194 phase A and B, respectively, both with 0.1% formic acid. The following linear gradient was
195 used: 1% B for 5 min; 1-5% B in 2 min; 5-35% B in 38 min; 35-50% B in 5 min; 50-90% B
196 in 5 min. Finally, the column was washed at 90% B for 10 min and then equilibrated at 1% B
197 for 20 min.
198 Peptide spectra were acquired in the 380−1800 *m*/*z* range at 30,000 resolution (full width at
199 half maximum, FWHM, at *m/z* 400) for the full scan. MS/MS spectra were acquired at 15,000
200 resolution (FWHM) in top 10 data-dependent acquisition (DDA) mode with rejection of
201 singly charged ions and of unassigned charge states. Precursors were fragmented by higher-
202 energy collisional dissociation (HCD) with 35% normalized collision energy and 2 *m*/*z*
203 isolation window. Dynamic exclusion was enabled with a repeat count of 1 and a repeat
204 duration of 30 s with exclusion duration of 20 s. For each sample, three technical replicates
205 were performed. Raw data files were acquired by Xcalibur software (version 2.2, Thermo
206 Fisher Scientific).
207

208 **2.4.2 Analysis of Short Peptides by Ultra-High Performance Liquid Chromatography-**
209 **MS/MS**

210 Samples were analyzed by RP chromatography as previously described (Montone, Capriotti,
211 et al., 2019) on a UHPLC Vanquish binary pump H coupled to a Q Exactive mass
212 spectrometer (Thermo Fisher Scientific, Bremen, Germany) by a heated electrospray (ESI)
213 source. Twenty μL of each sample were injected onto a Kinetex XB-C18 (100 × 2.1 mm, 2.6
214 μm particle size, Phenomenex, Torrance, USA) operated at 40 °C using Optima® LC-MS

215 grade H_2O (phase A) and ACN (phase B) both with 0.1% TFA (*v*/*v*) at 0.4 mL min^{-1}. The
216 chromatographic gradient was the following: 1% B for 2 min, 1–35% B in 20 min, 35–99% B
217 in 3 min; at the end of the gradient, a washing step at 99% B for 3 min and a re-equilibration
218 step at 1% B for 5 min were performed. The ESI source of the mass spectrometer was
219 operated as follows: 220 °C capillary temperature, 50 (arbitrary units, a.u.) sheath gas, 25
220 (a.u.) auxiliary gas, 0 (a.u.) sweep gas, 3200 V spray voltage, 280 °C auxiliary gas heater
221 temperature, 50 (%) S-Lens RF level. Full scan spectra were acquired in the positive
222 ionization mode in the range *m*/*z* 150–750 with a resolution (FWHM, at *m*/*z* 200) of 70,000
223 and automatic gain control (AGC) target value of 500,000 in full scan, max ion injection time
224 of 50 ms and the isolation window width was 2 *m*/*z*.
225 HCD MS/MS spectra acquisition was performed using top 5 DDA at 35% normalized
226 collision energy and 35,000 (FWHM, at *m*/*z* 200) resolution, AGC target value at 100,000
227 and dynamic exclusion at 3 s. Two inclusion lists with the exact *m*/*z* values for unique singly
228 charged precursor ions were used for DDA, one for non-modified peptides (4980 unique
229 masses) and one for peptides with modifications (3179 unique masses). Inclusion lists were
230 prepared using MatLab R2018, as previously described (Cerrato et al., 2020; Piovesana et al.,
231 2020; Piovesana, Capriotti, et al., 2019). Modifications were considered for 11 amino acids
232 side chains: hydroxyproline, hydroxylysine, methyllysine, dimethyllysine, trimethyllysine,
233 acetyllysine, succinyllysine, methylarginine, phosphoserine, methionine sulfoxide,
234 pyroglutamic acid. Lactic acid was also included as residue for building pseudo-peptide
235 combinations. Raw data files were acquired by Xcalibur software (version 3.1, Thermo
236 Fisher Scientific). Three technical replicates were performed for each sample, followed by a
237 blank sample.
238
239 **2.5. Peptide Identification**

240 **2.5.1. Medium-Sized Peptide Identification**

241 Peptides were identified by a combination of *de novo* spectra analysis and database search
242 against the Uniprot database for *Eupercaria incertae sedis* taxonomy (49444 sequences)
243 using PEAKS studio (version 7.5, Bioinformatics Solutions Inc., Waterloo, Canada) (Tran et
244 al., 2017). The following parameters were used: 10 ppm parent mass error, 0.02 Da fragment
245 mass error, no enzyme for digestion, 485 variable modifications with max 2 modifications for
246 each peptide, false discovery rate estimation option was enable and was always set to 1% at
247 PSM level, *de novo* score (ALC%) threshold was set at the maximum of 80% to leave only
248 the best quality identifications, peptide hit threshold (-10logP) was automatically calculated
249 by PEAKS to obtain the desired FDR of 1%.

250 The same raw files were also searched by the *de novo* freeware pNovo 3 (v 3.1.3, pFind
251 Team, Beijing, China) (Yang et al., 2019). Raw files were converted into .mgf file format
252 using ProteoWizard, applying the default settings. The following settings were used for
253 peptide identification by pNovo: 10 ppm and 20 ppm mass tolerance were set for precursor
254 and product ions, respectively; no enzyme specific digestion; the open search function was
255 set as false; top-1 function was used; the modifications available for search and selected as
256 variable were: acetyl[K]AnyN-term], acetyl[K], dimethyl[K], Gln>pyro-Gly[anyN-term],
257 glu>pyroGlu[anyN-term], methyl[K], methyl[R], oxidation[M], succinyl[K]. Results were
258 filtered based on PSM score, accepting for manual validation only peptides with a score >99.

259

260 **2.5.2. Short Peptide Identification**

261 Identification of short peptides was accomplished using the complete list of peptide masses
262 (with related sequence and molecular formula) and implementing it into Compound
263 Discoverer (v. 3.1, Thermo Fisher Scientific, Bremen, Germany).

Raw data files and a blank sample were processed using the workflow depicted in Figure S1 and with the settings shown in Table S1. Briefly, the workflow allowed to extract the masses from the raw data files according to customized parameters for *predict composition* tool (Table S1), align them, remove signals of the blank or lacking MS/MS and use the peptide lists to match the extracted features. Manual validation of the MS/MS spectra was also aided using the *compound class scoring* tool, which allowed to automatically match typical product ions deriving from amino acids at N-terminus, C-terminus and in the middle of the sequence, and assign them to 34 compound classes (20 for natural amino acids and 14 for modified amino acids; for the specific product ion details, refer to Table S2 and S3 for natural and modified amino acids, respectively). Manual validation of filtered features was aided by matching the experimental spectra to the in-silico spectra produced by mMass 3 (Strohalm et al., 2010). Peptides were tentatively identified according to the characteristic fragmentation spectra. For result comparison, raw files were also processed by pNovo.

3. Results and Discussion

In this study, both medium-sized and short peptides were addressed and identified from fish samples. This was done because the identification of endogenous peptides by conventional peptidomics approaches, which include SPE purification on C18, nanoHPLC coupled to high resolution MS and spectra matching to protein sequence databases using no enzyme specificity, is an appropriate method for large scale identification of peptides but it is not suitable for short sequences. In fact, this approach allows to identify peptides as short as 5 residue long, and needs to be complemented with a different strategy to account for shorter peptide sequences, which are particularly important, both because they are often responsible for bioactivity of peptides in most studies, and because they are likely to exist. In the case of fish, there is no clear specificity of protein cleavage (Ahmed et al., 2015). Calpains have a

289 preference for hydrophobic amino acids (mostly leucine, isoleucine and valine), for the
290 sequence Pro-Glu-Ser-Thr, and for calmodulin-binding domains. Their involvement in the
291 production of endogenous peptides in fish is in weakening the myofibrillar structure,
292 resulting in an increased accessibility to polypeptides by other proteases. Other enzymes
293 involved are lysosomal cathepsins, which are released from muscles after fish death. They
294 comprise 13 cathepsins, among which B, D and L play an important role in post-mortem
295 production of peptides from myofibrils, producing both large and smaller peptide fragments.
296 Cathepsin B has the capacity to degrade Z-Arg-Arg-MCA and Z-Phe-Arg-MCA, but also an
297 interesting peptidyledipeptidase activity, which results in the production of dipeptides from
298 the sequential cleavage of the C-terminal both of proteins and peptides. Cathepsin D has no
299 specificity but prefers cleavage at hydrophobic amino acid residues. Cathepsin L is highly
300 sensitive towards the synthetic substrate Z-Phe-Arg-MCA. Finally, other enzymes involved
301 in post-mortem peptide production are caspases, which are activated by the events that
302 initiate apoptosis and are the only enzymes mentioned so far to be considered specific, with
303 cleavage at Asp residues (Ahmed et al., 2015).
304 Given the complexity of the peptide mixture produced in fish muscle after fish death, a
305 comprehensive strategy for their characterization is here described, which allows to
306 specifically address short peptides and medium-size peptides. To provide a comprehensive
307 identification of native peptides in the fish muscle extract, PTMs were also considered. In this
308 sense, glutamic acid derivatives (such as the pyroglutamic residue), as well as N-lactoyl-
309 amino acids (pseudo-dipeptides) and succinyl-peptides represent possible targets in food
310 peptide characterization, as they are often reported to be taste-active amino acid derivatives,
311 often characterized by a pleasant taste (sweet, umami or kokumi), which could be exploited
312 to develop tastier, less bitter, and low-salt food products (Zhao et al., 2016). Though they
313 have been reported to be stable during gastrointestinal digestion, they also were not found to

314 have antihypertensive activity in dry-cured ham (Paolella et al., 2018). Other modifications
315 were proline and lysine hydroxylation, which is expected to derive from collagen. Finally,
316 acetylation, methylation, dimethylayion and trimetylation, methionine oxidation were
317 investigated as they have been recently found in potato hydrolysates and the decrease in
318 isoelectric point (pI) and increase in molecular weight of the peptides would potentially
319 influence the bioactivity of the same peptides, as well as the physico-chemical properties,
320 such as solubility in aqueous environment. As such, thought their abundance is expected to be
321 low, methods for their investigation and characterization are potentially useful to further mine
322 the bioactivity of food peptides and their organoleptic properties (Yao & Udenigwe, 2018).
323 Finally, phosphopeptides are known mineral binding peptides (Piovesana et al., 2018) and
324 were also included in the search.
325

326 **3.1. Characterization of Native Medium-Size Peptides in Fish**

327 The identification of medium-sized peptides was achieved by using an established
328 peptidomics approach. Currently, the use of techniques borrowed from shotgun proteomics
329 represents a reference field for the identification of peptides in a complex sample, as digests
330 or extracts with native peptides from food are. The challenges in this field for confident and
331 comprehensive peptide identification are provided by the limit in complete protein sequence
332 databases. For most organisms this is not an issue, while it still is in food characterization, as
333 in the case of this work. A complete protein sequence database for sea bass is currently not
334 available at time of this work, therefore proteins can be identified basically by similarity to
335 proteins reported for close organisms, to protein sequences obtained by translating
336 transcriptomic data or by *de novo* approaches, which do not rely on protein sequence
337 databases.

338 In this work, two strategies were employed and compared: the first one based on *de novo*
339 peptide identification, whereas the second one complements *de novo* with spectra matching to
340 protein sequence database. The use of *de novo* is the most advantageous strategy to identify
341 peptides from organisms that are not sequenced and for which the use of high taxonomical
342 entries results in a too large database, which in turn means too large search space, poor
343 peptide identification and very long computational times. This approach gives access to
344 peptide sequences not encrypted in the protein sequence database. In the first phase, the
345 freeware pNovo was used for identifying peptides by *de novo*. The large number of peptide
346 identifications provided by pNovo was manually filtered, based on the PSM score threshold
347 of 99 (which corresponds to 1% false discovery rate of peptide identification) and then each
348 spectrum manually checked using the graphic interface of the software. At this level, all
349 peptides, whose sequence was not obtained by extensive fragmentation of the chain with both
350 y- and b- ions, were excluded from the final list, as well as the peptides having spectra with
351 intense unmatched signals. For modified peptides, the same fragments used for validation of
352 short peptides were searched manually in the spectrum for confirmation. The strict criteria for
353 manual validation were necessary to reduce the number of false positive identifications. At
354 the end of this process, 119 peptide sequences were identified, with 17 of them having a
355 modification (1 pyroglutamic acid-peptide, 9 acetyllysine-peptides, 5 succinyllysine-peptides
356 and 1 methyllysine-peptide, Table S4). All peptides, except five, had lysine (50) or arginine
357 (64) as the last residue, which is compatible with carboxypeptidase B activity. Peptides were
358 6-16 amino acid long and were mainly hydrophilic (92% of sequences had a negative Grand
359 Average of Hydropathy (GRAVY) value) and basic (74% of the sequences had pI >8, Figure
360 S3). The most abundant residues, along with lysine and arginine, were isoleucine/leucine. In
361 this case, more than half of the residues were polar or with charged side chains (Figure S4).
362 The search for bioactive peptides was unsuccessful, and the use of PeptideRanker (Mooney et

al., 2012) to score the potential bioactivity indicated only one possible candidate with a score >0.8 (RRPPGWSPLR, Supporting Information, Table S4). The peptide showed partial sequences reported as ACE inhibitor sequences (GW, RP, RR) and DPP-VI inhibitor (GW, RP, RR, PP) after search of FeptideDB, a web-based utility which collects bioactive peptide data from both published research articles and available bioactive peptide databases (Panyayai et al., 2019).

To improve the number of peptide identifications, a complementary strategy was used, in which spectra identifications was complemented to database spectra matching to protein sequence databases using the PEAKS software. In this way, a large database was used (*Eupercaria incertae sedis* taxonomy, with 49444 sequences) and peptides, which might still be not represented in the database, were identified by *de novo*. Moreover, this strategy is the most effective for searching modifications. A total of 4077 peptides (with 2725 of them modified) were identified by database search and 2665 peptides (with 223 of them being modified) were identified by *de novo* only from all experiments (Supporting Information Table S5-7 and S8-10, respectively). Most peptides were uniquely identified by a single approach, with only 7 peptides common to both database and *de novo* by PEAKS. The number of common identifications with pNovo was high (56% of pNovo identifications).

Two-hundred seven different modifications were found for the peptides identified by database search, whereas the peptides identified by *de novo* had methionine oxidation as the only modification (Figure 1). The most represented modifications were mutations (829 peptides), as a result of amino acid substitutions in the protein sequence. This is no surprise, as there is no specific and complete database for sea bass proteome and variations can be expected by identification by similarity. The second most abundant modification was deamidation (of asparagine and glutamine, 194 peptides), followed by methionine oxidation (172 peptides), methyl ester (160 peptides). Thirty to 100 peptides were identified for

388 acetylation of protein N-term (92 peptides), carbamidomethylation (aspartic acid, histidine,
389 lysine or glutamic acid at protein N-term, 75 peptides), hydroxilation (68 peptides),
390 ethylation (67 peptides), dehydration (45 peptides), phosphorylation (43 peptides), lysine
391 acetylation (41 peptides), methylation (33 peptides), carbamylation (31 peptides). One
392 hundred ninety-four modifications had <30 peptides identifications (Figure 1, Supporting
393 Information Table S5-7).
394 As far as the modifications specifically addressed in this study were concerned, they were not
395 among the most abundant found, except for methionine oxidation. Hydroxylation was found
396 on lysine (17) and proline (8), but also on aspartic acid (23), arginine (14), asparagine and
397 tyrosine (3). Methylation was found on lysine (15 peptides) and arginine (2 peptides), but
398 also on serine (27 peptides), proline (6 peptides), histidine (2 peptides) and alanine,
399 asparagine, cysteine threonine (1 peptide). Dimethylation was not found on lysine, but on
400 other residues (6 peptides on arginine, 7 peptides on proline, 1 peptide on asparagine).
401 Trimethylation was found on lysine (3 peptides), but also on arginine (3 peptides) and alanine
402 (1 peptide). Acetyllysine was found in 41 peptides but also in 1 peptide as N-term
403 acetylation. Other acetylations were found at the peptide N-term for methionine (12
404 peptides), serine (11 peptides), alanine, aspartic acid (5 peptides), cysteine, glutamic acid (3
405 peptides), isoleucine, glycine (2 peptides), arginine, valine (1 peptide). Acetylation was
406 identified at the protein N-term for serine (126 peptides), methionine (23 peptides), alanine
407 (16 peptides), aspartic acid (6 peptides), glycine, phenylalanine (1 peptide). Finally,
408 acetylation was found for threonine (3 peptides). Succinylation was not found on lysine but
409 on cysteine (1 peptide), phosphoylation was found on serine (30 peptides), threonine (14
410 peptides) and tyrosine (6 peptides). No peptides modified with pyroglutamic acid were found,
411 despite being included in the search, whereas lactic acid is not a modification included in
412 protein PTMs. Lactic acid could have been introduced manually as a as a mass shift of

413 +72.0211 due to C(3)H(4)O(2) addition in the exact mass of precursor ions, with
414 modification on any N-term. Nevertheless, this was omitted, as lactoyl-derivatives are
415 generally pseudo-dipeptides (i.e., N-lactoyl-amino acids).

416 The physico-chemical feature analysis of the peptides identified by PEAKS indicated that
417 most peptides were hydrophilic, with negative GRAVY values (92% of the peptides
418 identified by database search and 89% of the peptides identified by *de novo*, Supporting
419 Information, Figures S5a and S6a, respectively). The molecular weight range was 598-6991
420 u, with modest differences between database and *de novo* peptides, with the latter being
421 mostly below 3000 u (93% vs 63% for peptides identified by database search, Supporting
422 Information, Figures S5b and S6b, respectively). No significant difference was found for the
423 pI distribution, with most peptides having pI > 8 (58% of the peptides identified by database
424 search and 59% of de novo peptides, Supporting Information, Figures S5c and S6c,
425 respectively). Results agreed with the ones obtained by pNovo search. Most peptides had
426 lysine or arginine as the last residue, which is compatible with carboxypeptidase B activity.

427 Both database and *de novo* peptide lists were searched for potential bioactivity by
428 PeptideRanker and PeptideDB. PeptideRanker provided 102 sequences with potential
429 bioactivity and score >80 for the peptides identified by database search; 117 sequences were
430 obtained for the peptides identified by *de novo* (Supporting Information Table S11 and S12).
431 The most represented bioactivities were ACE inhibitor and DPP-IV inhibitor activity, based
432 on recurrence of short bioactive peptides in the sequence of these peptides. No exact peptide
433 sequence was found bioactive.
434

435 **3.2. Characterization of Native Short Peptides in Fish**

436 Endogenous short peptide sequences, which for their molecular weights and ionization
437 properties resembles metabolites more than medium-sized peptides, were first cleaned up and

438 enriched on a GCB cartridge, whose strong absorption properties avoid losses during loading
439 and washing, unlike the common C18 cartridge employed for proteomics experiments. For
440 chromatographic peptide separation, a C18 stationary phase was chosen as it is the most used
441 in the works on bioactive peptides. The only main difference to conventional RP separation
442 of peptides to be employed for short sequences is TFA as mobile phase modifier, which is
443 necessary and cannot be substituted with formic acid. TFA is a strong ion-pairing agent that
444 induces the formation of heavier and more hydrophobic adducts, leading to a better retention
445 of the most hydrophilic sequences. C18 is not the only stationary phase useful for separation
446 of short peptides. In fact, based on our previous work (Piovesana, Capriotti, et al., 2019), RP
447 separation and zwitterionic hydrophilic interaction liquid chromatography (HILIC) turned out
448 to be comparable in terms of number of identified peptides, with the first allowing a higher
449 coverage of hydrophobic short peptides and the latter a higher coverage of hydrophilic
450 sequences. Therefore, based on the type of sample, the use of alternative stationary phases,
451 such as zwitterionic HILIC phases or porous graphitic carbon (Piovesana, Montone, et al.,
452 2019), may prove useful to improve the coverage of the low molecular weight peptidome and
453 provide complementary identifications to the conventional C18 used in this work.
454 To overcome the subsequent troubles in short peptide MS analysis, related to their low
455 endogenous abundance and poor ionization efficiency, which usually hinder the detection of
456 most sequences in standard metabolomics workflows, in this work a suspect screening
457 approach was exploited. More specifically, for most Orbitrap mass analyzers, DDA methods
458 are commonly preferred over data independent acquisition (DIA) approaches, because they
459 allow to acquire clearer and more diagnostic fragmentation spectra. However, many low-
460 abundance compounds are eventually not fragmented. Suspect screening approaches
461 constitute a valuable alternative for mass spectrometers which are too slow for DIA, because
462 low abundance compounds will still be fragmented if their *m/z* match with a list of expected

463 species inserted into the MS method. Amino acid combinations in di- tri- and tetrapeptides
464 were therefore obtained by means of Matlab, duplicate masses were filtered out and the
465 resulting list was implemented in the MS method. Suspect screening allowed, in fact, a 4-fold
466 increase in the number of identifications over the regular untargeted approach (Cerrato et al.,
467 2020; Piovesana, Capriotti, et al., 2019).
468 The aforementioned database was not merely employed for data acquisition, but also for data
469 processing as well. Compound Discoverer, in fact, was employed by virtue of a customized
470 workflow specifically developed for short peptide identification, which included the
471 comprehensive database as a *mass list* for the automatic match of experimental *m/z* to those
472 in the database (Cerrato et al., 2020; Piovesana, Capriotti, et al., 2019). The use of *mass list*,
473 as well as *predict composition* and *compound class scoring* tools allowed performing huge
474 filtering of false positive, with the result of a considerably faster and easier manual validation
475 of the sequences.
476 One hundred and fourteen short peptide sequences were eventually identified in the fish
477 extract, including 12 compounds presenting non-proteolytic aminoacyl derivatives (7 lactoyl-
478 amino acids, 2 pyroglutamic peptides, 1 hydroxyproline-peptide, 2 acetyllysine-peptides)
479 (Paolella et al., 2018) (Table S13). The proteinogenic amino acid sequences were mainly
480 tripeptides (45) and dipeptides (39), with only few tetrapeptides (18). Among the twelve
481 considered modifications, hydroxyproline, pyroglutamic acid, acetyllysine and lactic acid
482 containing sequences were identified, even though it is not sure whether such modified
483 residues were PTMs or artifacts generated after protein cleavage or during sample pre-
484 treatment. As far as the 102 short peptide sequences of proteinogenic amino acids were
485 concerned, they had a molecular weight in the range 189-525 u. Fifty peptides were
486 hydrophilic, with negative GRAVY values, and 52 were hydrophobic, with positive GRAVY
487 values, which span between -4/+4.5. Ninety-six short peptides had $pI<7$ (Figure S7).

488 The bioactivity of the found peptides was evaluated by searching FeptideDB and BIOPEP
489 database of bioactive peptides (Minkiewicz et al., 2019) and of sensory peptides (Iwaniak et
490 al., 2016).
491 For 39 peptide sequences, a matching to bioactive peptide databases was found for the exact
492 sequence (Figure 2). For additional 69 sequences a putative bioactivity was also assigned,
493 based on sequence similarity to bioactive peptides for which the bioactivity was previously
494 reported and that covered most of the sequence of the peptide identified in this study (Figure
495 2, Supporting information Table S13 for details of sequences and bioactivity assigned to each
496 short peptide). Bioactivity prediction based on similarity was possible because the function of
497 peptides is closely related to the amino acid sequence, and peptides with similar amino acid
498 sequences may exhibit similar bioactivity (Tu et al., 2018). Most peptides displayed more
499 than one bioactivity. The main bioactivities were the antihypertensive one as ACE inhibitors
500 (with renin inhibitor activity and vasoactive substance release activity also reported but as
501 minor ones) and the antidiabetic one as DPP-IV inhibitor (but also glucose uptake stimulation
502 activity and alfa-glucosidase inhibitor activity as a minor ones). These were the most
503 represented bioactivities for known and putative bioactive short peptides identified in this
504 work, probably because they are among the most investigated and reported in the databases
505 of bioactive peptides. The third most frequent bioactivity of native short peptides from fish
506 muscle was the dipeptidyl peptidase-III (DPP-III) inhibitor activity, which is expected to be
507 promising in pain management compounds (Khaket et al., 2015). Other minor bioactivities
508 were antioxidative and anxiolytic for known sequences and hypolipidemic and
509 antithrombotic, among the other sequences (Figure 2).
510 The investigation of sensory properties among the identified short peptides in BIOPEP
511 indicated that only a small amount of them has been reported as bitter (16), while two were
512 salty or umami peptides, while no data is reported for the others. This is potentially

513 interesting, as bitter taste is one of the main drawbacks in the application of bioactive protein
514 hydrolysates and the production of functional foods (Hajfathalian et al., 2019).
515 As far as the modified peptides were concerned, the dipeptide Phe-Hyp was previously
516 reported as one of the peptides released from collagen hydrolysis. Though no activity was
517 evaluated on this specific sequence, Pro-Hyp and Hyp-Gly were examined by using tissue
518 and cell culture systems, demonstrating that they can enhance the growth of fibroblasts and
519 the production of hyaluronic acid from human dermal fibroblasts, which has been associated
520 with beneficial effects on skin and joints (Sato et al., 2019).
521 As far as the amino acid residues occurrence was concerned, most of them were hydrophobic
522 (mostly phenylalanine, leucine or isoleucine, valine), followed by acid (glutamic acid) and
523 polar (tyrosine), with only few having basic side chains (Figure S8). The distribution allowed
524 to further mine the potential bioactivity of this extract, because it is related to the presence of
525 certain key amino acids. Hydrophobic amino acids, such as valine and proline, are very
526 important in most antihypertensive peptides (Tu et al., 2018) and in DPP-IV inhibitor activity
527 of peptides, which agrees with the most common bioactivity reported by searching BIOPEP.
528 Similarly, aromatic residues, but also histidine, cysteine, proline and methionine, have been
529 linked to the antioxidant activity of food, and they are also represented in these peptides,
530 which are particularly rich in aromatic residues (mostly phenylalanine). The large presence of
531 hydrophobic amino acids, especially at the C-terminus, was compatible with hydrolysis by
532 calpains and cathepsin D. The large presence of dipeptides is compatible with the action of
533 cathepsin B.
534 In order to prove the viability of our approach, pNovo was also tested for *de novo* sequencing
535 of short peptides. De novo approach represents the only viable option for comparing our
536 developed methodology to already existing techniques, since common proteomics software,
537 which are programmed for protein identification, fail in the identification of short peptides,

538 because the sequence can be attributed to too many proteins and result in extremely low
539 levels of confidence, therefore peptides shorter than 5 amino acids are usually not considered
540 in these approaches. However, only 10 short peptides were identified by means of pNovo,
541 which were all common to the ones identified by Compound Discoverer. Moreover, the order
542 of the amino acids was sometimes misinterpreted, because peak intensity of product ions is
543 not considered in this software. However, relative intensities of the product ions are often
544 useful during manual validation for determining the correct amino acid sequence. The
545 unsatisfactory results were both caused by the lower cut off at *m/z* 300, which excludes most
546 dipeptides, and the peculiar fragmentation pathways of short peptides, which include many
547 low *m/z* range product ions, which are usually neglected in the study of medium-sized
548 peptides spectra. Another drawback of this approach was the long list of candidate peptides to
549 be manually validated, due to the software not integrating peaks but only extracting features
550 from raw data files. While *de novo* strategies can be successfully applied to the identification
551 of medium-sized peptides in organisms with unknown genomes, their applications for the
552 characterization of the short peptidome appear to date rather limited if de novo software will
553 not be suitably modified for the specific needs of short peptide sequences. At the moment,
554 metabolomics approaches, in which great attention is given to peak extraction, alignment and
555 manual validation of the MS/MS spectra, have much greater potential, especially when
556 suspect screening approaches are chosen. Those considerations clearly indicate that
557 endogenous short peptides data analysis resembles that of metabolites more than that of
558 medium-sized peptide sequences.
559

560 **4. Conclusions**

561 The strategy described in this work, and applied to sea bass native peptide extract, allows to
562 obtain a comprehensive characterization of native peptides in a complex sample. Two clean-

563 up strategies were applied in parallel to allow mining both medium-size and short peptide
564 sequences. In both cases, modifications were considered, as they are underexplored but
565 potentially interesting for bioactivity and taste effects. Medium-size peptides were identified
566 by *de novo* and combination of *de novo* and database spectra matching to protein sequence
567 database, to increase the number of peptides identified in the study and better unravel
568 modifications possible present. The latter approach proved successful and provided very large
569 number of identified peptides and more than 200 different modifications. For short peptides,
570 a dedicated workflow was developed, which included not only clean-up, but also MS
571 acquisition and raw data file analysis. A suspect screening-based strategy was used for
572 spectra acquisition and Compound Discoverer was then exploited for data analysis and speed
573 up the process of peptide identification. More than 100 short peptides were identified, most of
574 them were found bioactive or related to taste. As such, the developed method provides a
575 comprehensive characterization of complex peptide mixtures and can be applied to any type
576 of sample in bioactive peptide research.

577

578 **5. References**

579 Ahmed, Z., Donkor, O., Street, W. A., & Vasiljevic, T. (2015). Calpains- and cathepsins-
580 induced myofibrillar changes in post-mortem fish: Impact on structural softening and
581 release of bioactive peptides. *Trends in Food Science and Technology*, *45*, 130–146.
582 https://doi.org/10.1016/j.tifs.2015.04.002
583 Capriotti, A. L., Caruso, G., Cavaliere, C., Samperi, R., Ventura, S., Zenezini Chiozzi, R., &
584 Laganà, A. (2015). Identification of potential bioactive peptides generated by simulated
585 gastrointestinal digestion of soybean seeds and soy milk proteins. *Journal of Food*
586 *Composition and Analysis*, *44*, 205–213. https://doi.org/10.1016/j.jfca.2015.08.007
587 Capriotti, A. L., Cavaliere, C., Foglia, P., Piovesana, S., Samperi, R., Zenezini Chiozzi, R., &

588 Laganà, A. (2015). Development of an analytical strategy for the identification of
589 potential bioactive peptides generated by in vitro tryptic digestion of fish muscle
590 proteins. *Analytical and Bioanalytical Chemistry*, *407*(3), 845–854.
591 https://doi.org/10.1007/s00216-014-8094-z
592 Capriotti, A. L., Cavaliere, C., Piovesana, S., Samperi, R., & Laganà, A. (2016). Recent
593 trends in the analysis of bioactive peptides in milk and dairy products. *Analytical and*
594 *Bioanalytical Chemistry*, *408*(11), 2677–2685. https://doi.org/10.1007/s00216-016-
595 9303-8
596 Castellano, P., Mora, L., Escudero, E., Vignolo, G., Aznar, R., & Toldrá, F. (2016).
597 Antilisterial peptides from Spanish dry-cured hams: Purification and identification. *Food*
598 *Microbiology*, *59*, 133–141. https://doi.org/10.1016/j.fm.2016.05.018
599 Cerrato, A., Aita, S. E., Capriotti, A. L., Cavaliere, C., Montone, C. M., Laganà, A., &
600 Piovesana, S. (2020). A new opening for the tricky untargeted investigation of natural
601 and modified short peptides. *Talanta*, *in press*.
602 Gallego, M., Aristoy, M. C., & Toldrá, F. (2014). Dipeptidyl peptidase IV inhibitory peptides
603 generated in Spanish dry-cured ham. *Meat Science*, *96*, 757–761.
604 https://doi.org/10.1016/j.meatsci.2013.09.014
605 Hajfathalian, M., Ghelichi, S., García-Moreno, P. J., Moltke Sørensen, A. D., & Jacobsen, C.
606 (2019). Peptides: Production, bioactivity, functionality, and applications. *Critical*
607 *Reviews in Food Science and Nutrition*, *58*(18), 3097–3129.
608 https://doi.org/10.1080/10408398.2017.1352564
609 Iwaniak, A., Minkiewicz, P., Darewicz, M., Sieniawski, K., & Starowicz, P. (2016). BIOPEP
610 database of sensory peptides and amino acids. *Food Research International*, *85*, 156–
611 161. https://doi.org/10.1016/j.foodres.2016.04.031
612 Khaket, T. P., Redhu, D., Dhanda, S., & Singh, J. (2015). In silico evaluation of potential

613 DPP-III inhibitor precursors from dietary proteins. *International Journal of Food*
614 *Properties*, *18*, 499–507. https://doi.org/10.1080/10942912.2013.787626
615 Liu, D., Chen, X., Huang, M., & Zhou, G. (2019). Antioxidant activity of peptides in
616 postmortem aged duck meat as affected by cooking and in vitro digestion. *International*
617 *Journal of Food Properties*, *22*(1), 727–736.
618 https://doi.org/10.1080/10942912.2019.1605374
619 Liu, R., Xing, L., Fu, Q., Zhou, G. H., & Zhang, W. G. (2016). A review of antioxidant
620 peptides derived from meat muscle and by-products. *Antioxidants*, *5*(3), 32.
621 https://doi.org/10.3390/antiox5030032
622 Minkiewicz, P., Iwaniak, A., & Darewicz, M. (2019). BIOPEP-UWM database of bioactive
623 peptides: Current opportunities. *International Journal of Molecular Sciences*, *20*(23),
624 5978. https://doi.org/10.3390/ijms20235978
625 Montone, C. M., Capriotti, A. L., Cavaliere, C., La Barbera, G., Piovesana, S., Zenezini
626 Chiozzi, R., & Laganà, A. (2018a). Characterization of antioxidant and angiotensin-
627 converting enzyme inhibitory peptides derived from cauliflower by-products by
628 multidimensional liquid chromatography and bioinformatics. *Journal of Functional*
629 *Foods*, *44*, 40–47. https://doi.org/10.1016/j.jff.2018.02.022
630 Montone, C. M., Capriotti, A. L., Cavaliere, C., La Barbera, G., Piovesana, S., Zenezini
631 Chiozzi, R., & Laganà, A. (2018b). Peptidomic strategy for purification and
632 identification of potential ACE-inhibitory and antioxidant peptides in Tetradesmus
633 obliquus microalgae. *Analytical and Bioanalytical Chemistry*, *410*(15), 3573–3586.
634 https://doi.org/10.1007/s00216-018-0925-x
635 Montone, C. M., Capriotti, A. L., Cerrato, A., Antonelli, M., La Barbera, G., Piovesana, S.,
636 Laganà, A., & Cavaliere, C. (2019). Identification of bioactive short peptides in cow
637 milk by high-performance liquid chromatography on C18 and porous graphitic carbon

638 coupled to high-resolution mass spectrometry. *Analytical and Bioanalytical Chemistry*,
639 *411*(15), 3395–3404. https://doi.org/10.1007/s00216-019-01815-0
640 Montone, C. M., Zenezini Chiozzi, R., Marchetti, N., Cerrato, A., Antonelli, M., Capriotti, A.
641 L., Cavaliere, C., Piovesana, S., & Laganà, A. (2019). Peptidomic approach for the
642 identification of peptides with potential antioxidant and anti-hyperthensive effects
643 derived from Asparagus by-products. *Molecules*, *24*(19), 3627.
644 https://doi.org/10.3390/molecules24193627
645 Mooney, C., Haslam, N. J., Pollastri, G., & Shields, D. C. (2012). Towards the Improved
646 Discovery and Design of Functional Peptides: Common Features of Diverse Classes
647 Permit Generalized Prediction of Bioactivity. *PLoS ONE*, *7*(10), e45012.
648 https://doi.org/10.1371/journal.pone.0045012
649 Mora, L., Escudero, E., Fraser, P. D., Aristoy, M. C., & Toldrá, F. (2014). Proteomic
650 identification of antioxidant peptides from 400 to 2500Da generated in Spanish dry-
651 cured ham contained in a size-exclusion chromatography fraction. *Food Research*
652 *International*, *56*, 68–76. https://doi.org/10.1016/j.foodres.2013.12.001
653 Mora, L., Gallego, M., Reig, M., & Toldrá, F. (2017). Challenges in the quantitation of
654 naturally generated bioactive peptides in processed meats. *Trends in Food Science and*
655 *Technology*, *69*, 306–314. https://doi.org/10.1016/j.tifs.2017.04.011
656 Mora, L., González-Rogel, D., Heres, A., & Toldrá, F. (2020). Iberian dry-cured ham as a
657 potential source of α-glucosidase-inhibitory peptides. *Journal of Functional Foods*, *67*,
658 103840. https://doi.org/10.1016/j.jff.2020.103840
659 Nongonierma, A. B., & FitzGerald, R. J. (2016). Strategies for the discovery, identification
660 and validation of milk protein-derived bioactive peptides. *Trends in Food Science and*
661 *Technology*, *50*, 26–43. https://doi.org/10.1016/j.tifs.2016.01.022
662 Nurdiani, R., Vasiljevic, T., Yeager, T., Singh, T. K., & Donkor, O. N. (2017). Bioactive

663 peptides with radical scavenging and cancer cell cytotoxic activities derived from
664 Flathead (Platycephalus fuscus) by-products. *European Food Research and Technology*,
665 *243*, 627–637. https://doi.org/10.1007/s00217-016-2776-z
666 O'Keeffe, M. B., & Fitzgerald, R. J. (2015). Identification of short peptide sequences in
667 complex milk protein hydrolysates. *Food Chemistry*, *184*, 140–146.
668 https://doi.org/10.1016/j.foodchem.2015.03.077
669 Panyayai, T., Ngamphiw, C., Tongsima, S., Mhuantong, W., Limsripraphan, W.,
670 Choowongkomon, K., & Sawatdichaikul, O. (2019). FeptideDB: A web application for
671 new bioactive peptides from food protein. *Heliyon*, *5*(7), e02076.
672 https://doi.org/10.1016/j.heliyon.2019.e02076
673 Paolella, S., Prandi, B., Falavigna, C., Buhler, S., Dossena, A., Sforza, S., & Galaverna, G.
674 (2018). Occurrence of non-proteolytic amino acyl derivatives in dry-cured ham. *Food*
675 *Research International*, *114*, 38–46. https://doi.org/10.1016/j.foodres.2018.07.057
676 Piovesana, S., Capriotti, A. L., Cavaliere, C., La Barbera, G., Montone, C. M., Zenezini
677 Chiozzi, R., & Laganà, A. (2018). Recent trends and analytical challenges in plant
678 bioactive peptide separation, identification and validation. *Analytical and Bioanalytical*
679 *Chemistry*, *410*(15), 3425–3444. https://doi.org/10.1007/s00216-018-0852-x
680 Piovesana, S., Capriotti, A. L., Cerrato, A., Crescenzi, C., La Barbera, G., Laganà, A.,
681 Montone, C. M., & Cavaliere, C. (2019). Graphitized Carbon Black Enrichment and
682 UHPLC-MS/MS Allow to Meet the Challenge of Small Chain Peptidomics in Urine.
683 *Analytical Chemistry*, *91*(17), 11474–11481.
684 https://doi.org/10.1021/acs.analchem.9b03034
685 Piovesana, S., Cerrato, A., Antonelli, M., Benedetti, B., Capriotti, A. L., Cavaliere, C.,
686 Montone, C. M., & Laganà, A. (2020). A clean-up strategy for identification of
687 circulating endogenous short peptides in human plasma by zwitterionic hydrophilic

688 liquid chromatography and untargeted peptidomics identification. *Journal of*
689 *Chromatography A*, *1613*, 460699. https://doi.org/10.1016/j.chroma.2019.460699
690 Piovesana, S., Montone, C. M., Cavaliere, C., Crescenzi, C., La Barbera, G., Laganà, A., &
691 Capriotti, A. L. (2019). Sensitive untargeted identification of short hydrophilic peptides
692 by high performance liquid chromatography on porous graphitic carbon coupled to high
693 resolution mass spectrometry. *Journal of Chromatography A*, *1590*, 73–79.
694 https://doi.org/10.1016/j.chroma.2018.12.066
695 Sato, K., Jimi, S., & Kusubata, M. (2019). Generation of bioactive prolyl-hydroxyproline
696 (Pro-Hyp) by oral administration of collagen hydrolysate and degradation of endogenous
697 collagen. *International Journal of Food Science and Technology*, *54*, 1976–1980.
698 https://doi.org/10.1111/ijfs.14145
699 Strohalm, M., Kavan, D., Novák, P., Volný, M., & Havlíček, V. (2010). mMass 3: A Cross-
700 Platform Software Environment for Precise Analysis of Mass Spectrometric Data.
701 *Analytical Chemistry*, *82*(11), 4648–4651. https://doi.org/10.1021/ac100818g
702 Tran, N. H., Zhang, X., Xin, L., Shan, B., & Li, M. (2017). De novo peptide sequencing by
703 deep learning. *Proceedings of the National Academy of Sciences of the United States of*
704 *America*, *114*(31), 8247–8252. https://doi.org/10.1073/pnas.1705691114
705 Tu, M., Cheng, S., Lu, W., & Du, M. (2018). Advancement and prospects of bioinformatics
706 analysis for studying bioactive peptides from food-derived protein: Sequence, structure,
707 and functions. *TrAC - Trends in Analytical Chemistry*, *105*, 7–17.
708 https://doi.org/10.1016/j.trac.2018.04.005
709 Valero, Y., Saraiva-Fraga, M., Costas, B., & Guardiola, F. A. (2020). Antimicrobial peptides
710 from fish: beyond the fight against pathogens. *Reviews in Aquaculture*, *12*, 224–253.
711 https://doi.org/10.1111/raq.12314
712 Xing, L. J., Hu, Y. Y., Hu, H. Y., Ge, Q. F., Zhou, G. H., & Zhang, W. G. (2016).

713 Purification and identification of antioxidative peptides from dry-cured Xuanwei ham. *Food Chemistry*, *194*, 951–958. https://doi.org/10.1016/j.foodchem.2015.08.101
714
715 Yang, H., Chi, H., Zeng, W. F., Zhou, W. J., & He, S. M. (2019). pNovo 3: Precise de novo
716 peptide sequencing using a learning-to-rank framework. *Bioinformatics*, *35*(14), i183–
717 i190. https://doi.org/10.1093/bioinformatics/btz366
718 Yao, S., & Udenigwe, C. C. (2018). Peptidomics of potato protein hydrolysates: Implications
719 of post-translational modifications in food peptide structure and behaviour. *Royal*
720 *Society Open Science*, *5*(7), 172425. https://doi.org/10.1098/rsos.172425
721 Yathisha, U. G., Bhat, I., Karunasagar, I., & Mamatha, B. S. (2019). Antihypertensive
722 activity of fish protein hydrolysates and its peptides. *Critical Reviews in Food Science*
723 *and Nutrition*, *59*(15), 2363–2374. https://doi.org/10.1080/10408398.2018.1452182
724 Zenezini Chiozzi, R., Capriotti, A. L., Cavaliere, C., La Barbera, G., Montone, C. M.,
725 Piovesana, S., & Laganà, A. (2018). Label-Free Shotgun Proteomics Approach to
726 Characterize Muscle Tissue from Farmed and Wild European Sea Bass (Dicentrarchus
727 labrax). *Food Analytical Methods*, *11*(1), 292–301. https://doi.org/10.1007/s12161-017-
728 0999-7
729 Zenezini Chiozzi, R., Capriotti, A. L., Cavaliere, C., La Barbera, G., Piovesana, S., Samperi,
730 R., & Laganà, A. (2016). Purification and identification of endogenous antioxidant and
731 ACE-inhibitory peptides from donkey milk by multidimensional liquid chromatography
732 and nanoHPLC-high resolution mass spectrometry. *Analytical and Bioanalytical*
733 *Chemistry*, *408*(20), 5657–5666. https://doi.org/10.1007/s00216-016-9672-z
734 Zhao, C. J., Schieber, A., & Gänzle, M. G. (2016). Formation of taste-active amino acids,
735 amino acid derivatives and peptides in food fermentations – A review. *Food Research*
736 *International*, *89*, 39–47. https://doi.org/10.1016/j.foodres.2016.08.042
737

738 **Figure caption**

739 **Figure 1.** number of modified peptides identified by PEAKS for database search (blue bars)
740 and *de novo* (red bar) from sea bass native peptides. Only modifications with >30 peptides are
741 reported.

742

743 **Figure 2**. number of peptides identified in the characterization of short native peptides in fish
744 muscle and displaying a reported bioactivity (blue) or for a which a bioactivity could be
745 putatively assigned based on sequence coverage with reported bioactive peptide sequences
746 (orange).

Figure 1
Click here to download high resolution image

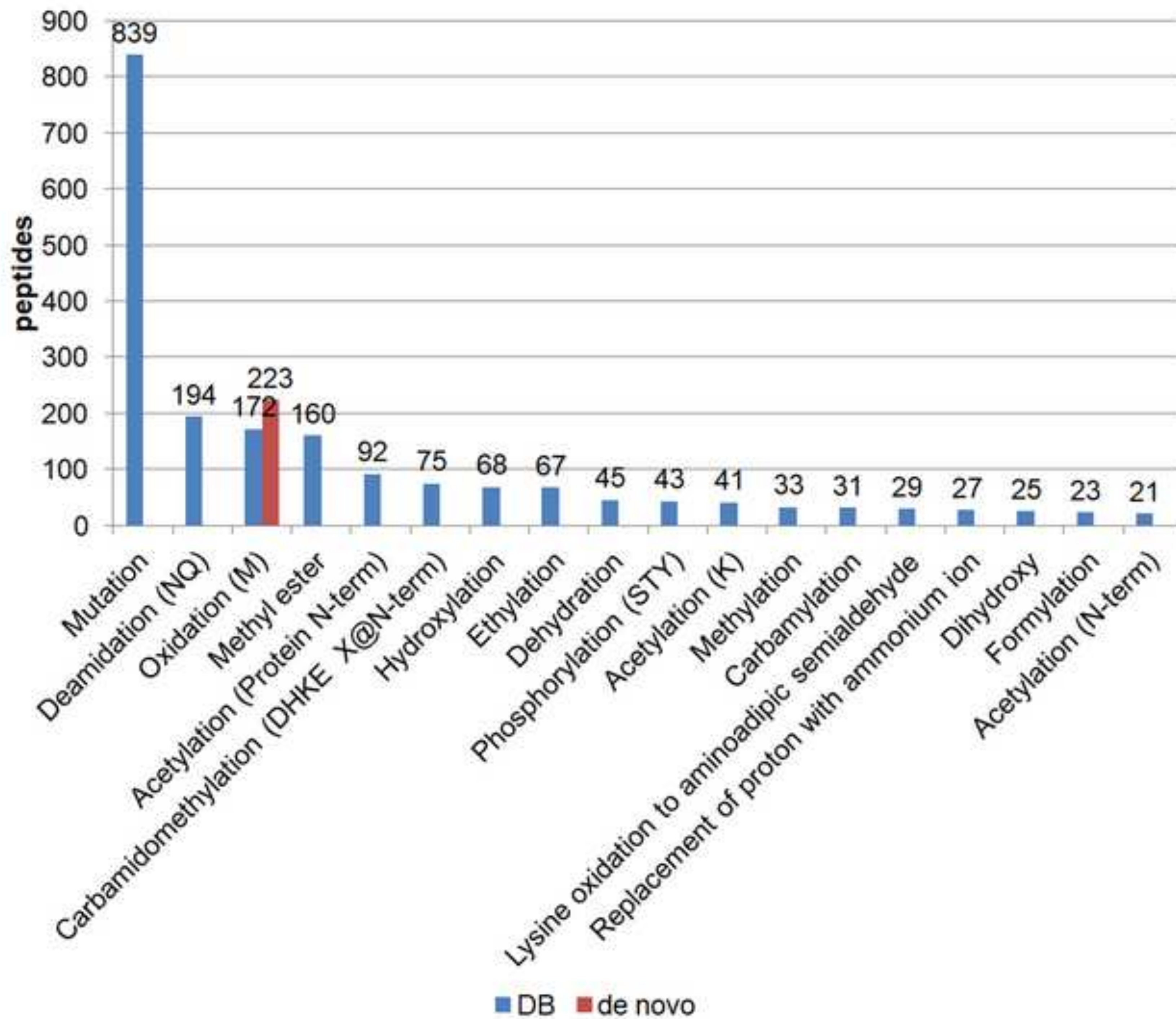
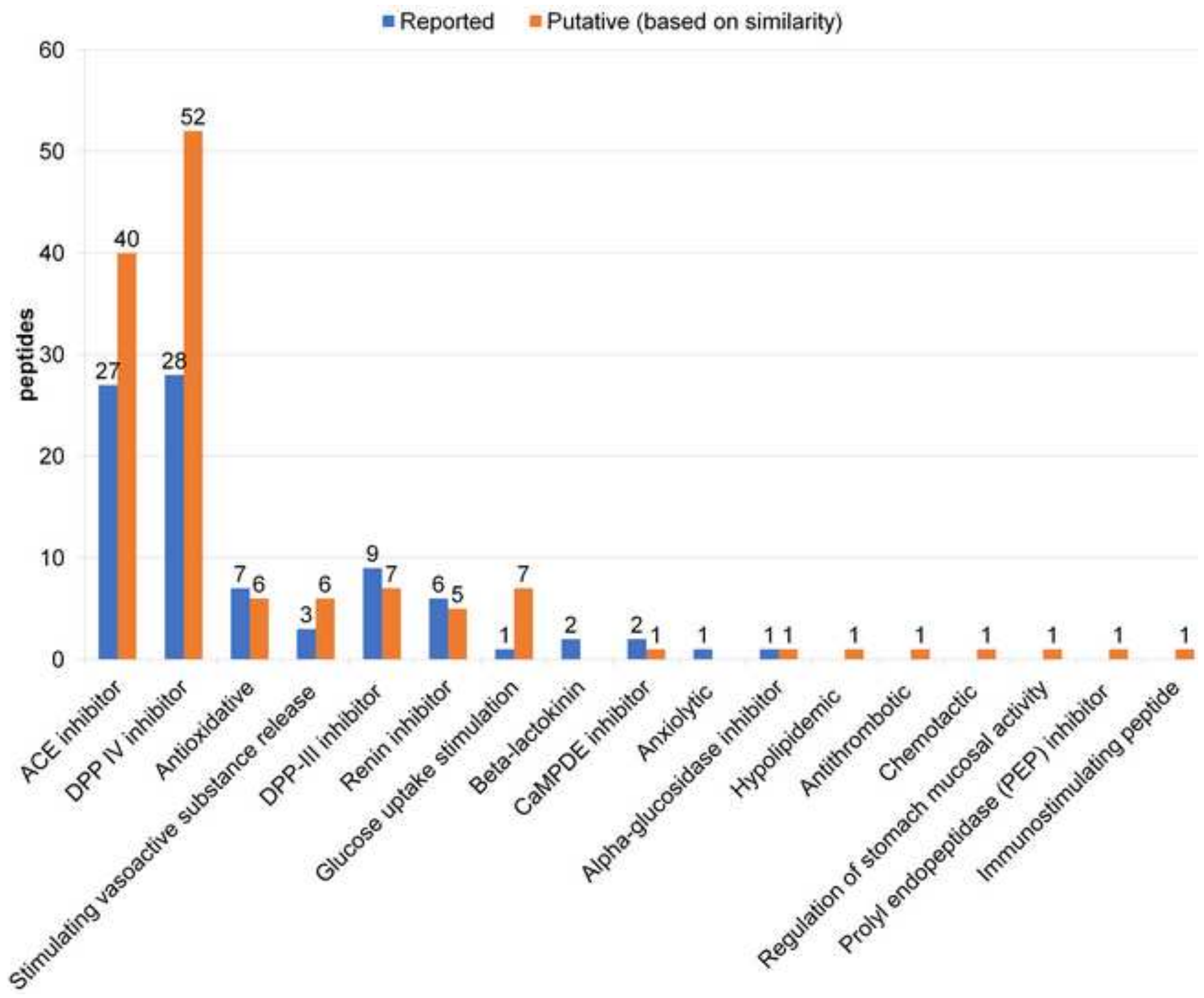


Figure 2

Click here to download high resolution image



Supplementary Material
Click here to download Supplementary Material: Supporting Information.docx

Supporting Information

Comprehensive Identification of Native Medium-Sized and Short Bioactive Peptides in Sea Bass Muscle

Andrea Cerrato[1], Sara Elsa Aita[1], Chiara Cavaliere[1], Aldo Laganà[1,2], Carmela Maria Montone[1], Susy Piovesana*[1], Riccardo Zenezini Chiozzi[3,4], Anna Laura Capriotti[1]

1. Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy
2. CNR NANOTEC, Campus Ecotekne, University of Salento, Via Monteroni, 73100 Lecce, Italy
3. Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

4. Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

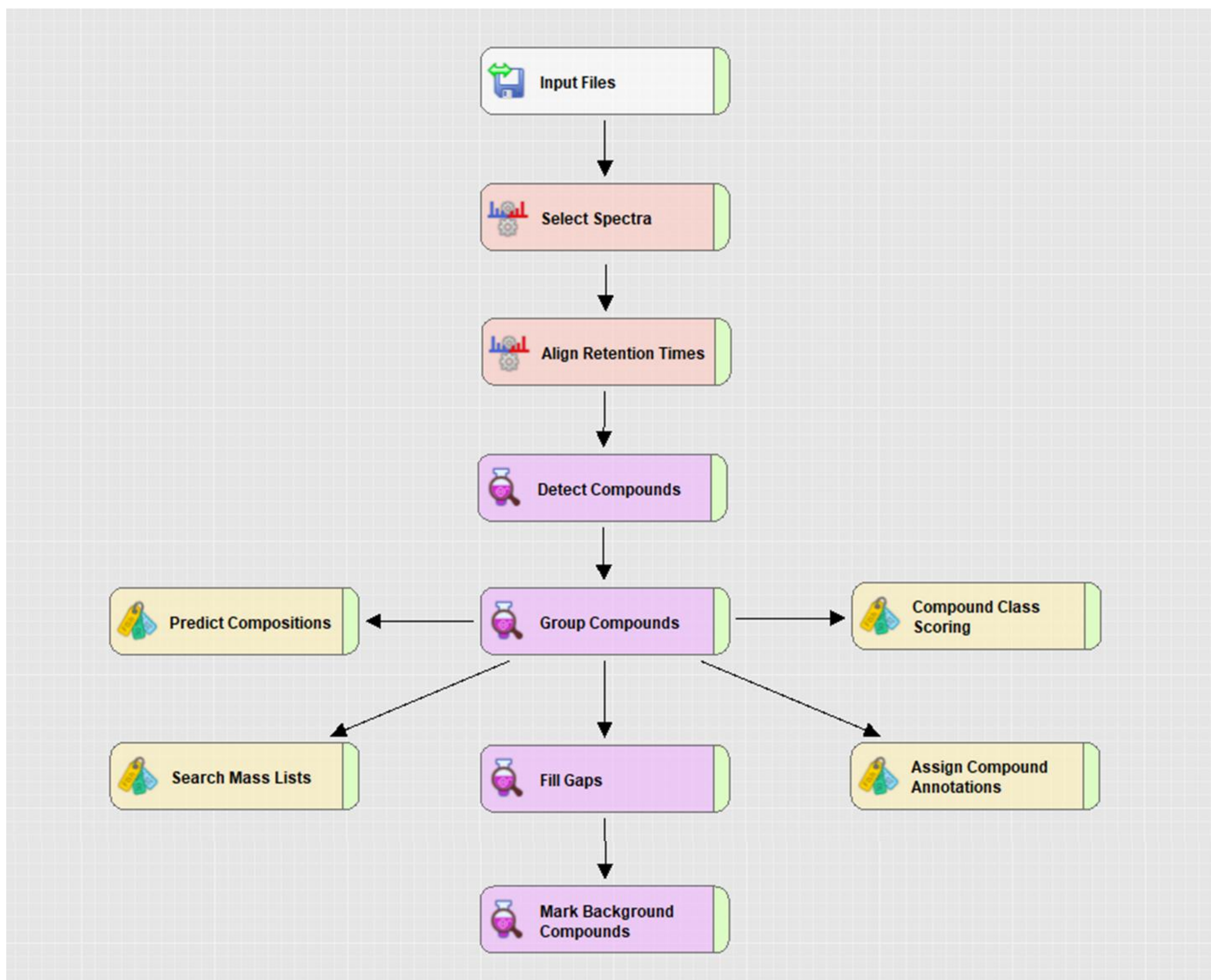


Figure S1. Data processing workflow for short peptide identification on Compound Discoverer

Table S1. Detailed Compound Discoverer parameters for short peptide data processing

Select spectra	
Precursor Selection	Use MS(n-1) Precursor
Use Isotope Pattern in Precursor Reevaluation	True
Provide Profile Spectra	Automatic
Lower RT Limit	0
Upper RT Limit	25
First Scan	0
Min Precursor Mass	0 Da
Max Precursor Mass	750 Da
Minimum Peak Count	1
MS Order	Any
Polarity mode	Positive
S/N Threshold	1.5
Align Retention Time	
Alignment Model	Adaptive Curve
Alignment Fallback	Use Linear Model
Minimum Shift [min]	0.2
Shift Reference File	Ture
Mass Tolerance	5 ppm
Remove Outlier	True
Detect Compounds	
Mass Tolerance [ppm]	5 ppm
Intensity Tolerance [%]	30
S/N Threshold	3
Min. Peak Intensity	100000
Ions (positive mode)	[M + H]+1
Base Ions	[M + H]+1
Min. Element Counts	C H
Max. Element Counts	C50 H60 O15 N15
Filter Peaks	True
Max. Peak Width [min]	0.5
Min #Scans per Peak	5
Min #Isotopes	1
Group Compounds	
Mass Tolerance	5 ppm

RT Tolerance	0.2
Preferred Ions	[M+H]+1
Predict Composition	
Mass Tolerance	5 ppm
Min. Element Counts	C H
Max. Elements Counts	C50 H60 O15 N15
Min. RDBE	0
Max. RDBE	30
Min. H/C	0.1
Max. H/C	4
Intensity Tolerance [%]	30
Intensity Threshold [%]	0.1
S/N Threshold	3
Min. Spectral Fit [%]	30
Min. Pattern Cov. [%]	90
Use Dynamic Recalibration	True
Search Mass List	
Use Retention Time	False
Mass Tolerance	5 ppm
Compound Class Scoring	
S/N Threshold	10
High Acc. Mass Tolerance	10 ppm
Allow AIF Scoring	True
Fill Gaps	
Mass Tolerance	5 ppm
S/N Threshold	1.5
Use Real Peak Detection	True
Mark Background Compounds	
Max. Sample/Blank	5
Hide background	True

Table S2. Product ions deriving from natural amino acid fragmentations. Fragments in green were inserted in compound class scoring tool based on experimental data or spectra reported in the literature. Fragments in red would have been inserted since were found to be typical but were omitted because they fall below the cut off at *m/z* 50 for MS/MS fragmentation

	Ala	**Arg**	**Asn**	**Asp**	**Cys**
Formula	$C_3H_7NO_2$	$C_6H_{14}N_4O_2$	$C_4H_8N_2O_3$	$C_4H_7NO_4$	$C_3H_7NO_2S$
Molecular Weight	89.047679	174.111676	132.053493	133.037509	121.019751
$[M+H]^+$	90.054955	175.118952	133.060769	134.044785	122.027027
Iminium Ion (*a* ion)	44.0495	**129.1135**	**87.0553**	**88.0393**	**76.0215**
Im – H₂O	-	-	-	**70.0287**	-
Im – NH₃	-	**112.0869**	**70.0287**	-	-
b ion	**72.0444**	157.1084	**115.0502**	**116.0342**	**104.0165**
b – H₂O	-	-	-	98.0237	-
b – NH₃	55.0178	**140.0818**	98.0237	99.0077	86.9899
c ion	89.0709	174.1349	132.0768	133.0608	121.0430
x ion	116.0342	201.0982	159.0400	160.0241	148.0063
y ion	**90.0550**	**175.1190**	**133.0608**	**134.0448**	**122.0270**
y – H₂O	-	-	-	**116.0342**	-
y – NH₃ (z ion)	-	**158.0924**	**116.0342**	-	-
Others	-	**70.0561**; **116.0706**	-	-	-

	Gly	**Glu**	**Gln**	**His**	**Xle**
Formula	$C_2H_5NO_2$	$C_5H_9NO_4$	$C_5H_9N_2O_3$	$C_6H_9N_3O_2$	$C_6H_{13}NO_2$
Molecular Weight	75.032029	147.053159	132.053493	133.037509	121.019751
$[M+H]^+$	76.039305	148.060435	133.060769	134.044785	122.027027
Iminium Ion (*a* ion)	30.0338	**102.0550**	**101.0709**	**110.0713**	**86.0964**
Im – H₂O	-	**84.0444**	-	-	-
Im – NH₃	-	-	**84.0444**	-	69.0699
b ion	**58.0287**	**130.0499**	**129.0659**	**138.0662**	**114.0913**
b – H₂O	-	112.0393	-	-	-
b – NH₃	-	113.0233	112.0393	121.0396	97.0648

c ion	75.0533	147.0764	146.0924	155.0927	131.1179
x ion	102.0186	174.0397	173.0557	182.0560	158.0812
y ion	**76.0393**	**148.0604**	**147.0764**	**156.0768**	**132.1019**
y – H_2O	-	**130.0499**	-	-	-
y – NH_3 (z ion)	-	-	**130.0499**	-	-
Others	-	-	-	-	-
	Lys	**Met**	**Phe**	**Pro**	**Ser**
Formula	$C_6H_{14}N_2O_2$	$C_5H_{11}NO_2S$	$C_9H_{11}NO_2$	$C_5H_7NO_2$	$C_3H_7NO_3$
Molecular Weight	146.105528	149.051051	165.078979	115.063329	105.042594
$[M+H]^+$	147.112804	150.058327	166.086255	116.070605	106.049870
Iminium Ion (*a* ion)	101.1073	**104.0528**	**120.0808**	**70.0651**	**60.0444**
Im – H_2O	-	-	-	-	42.0338
Im – NH_3	**84.0808**	-	103.0542	-	-
b ion	**129.1022**	**132.0478**	**148.0757**	**98.0600**	**88.0393**
b – H_2O	-	-	-	98.0237	-
b – NH_3	112.0757	-	-	99.0077	86.9899
c ion	146.1288	149.0743	165.1022	**115.0866**	105.0659
x ion	173.0921	176.0376	192.0655	142.0499	132.0291
y ion	**147.1128**	**150.0583**	**166.0863**	**116.0706**	**106.0499**
y – H_2O	129.1022	-	-	**116.0342**	88.0393
y – NH_3 (z ion)	**130.0863**	**133.0318**	**149.0597**	-	-
Others	-	-	**131.0941**	-	-
	Thr	**Trp**	**Tyr**	**Val**	
Formula	$C_4H_9NO_3$	$C_{11}H_{12}N_2O_2$	$C_9H_{11}NO_3$	$C_5H_{11}NO_2$	
Molecular Weight	119.058244	204.089878	181.073894	117.078979	
$[M+H]^+$	120.06552	205.097154	182.081170	118.086255	
Iminium Ion (*a* ion)	**74.0600**	**159.0917**	**136.0757**	**72.0808**	
Im – H_2O	**56.0495**	-	-	-	
Im – NH_3	-	-	-	55.0542	

b ion	**102.0550**	**187.0866**	**164.0706**	**100.0757**
b – H_2O	84.0444	-	146.0600	-
b – NH_3	-	-	-	-
c ion	119.0815	204.1131	181.0972	117.1022
x ion	146.0448	231.0764	208.0604	144.0655
y ion	**120.0655**	**205.0972**	**182.0812**	**118.0863**
y – H_2O	**102.0550**	-	-	-
y – NH_3 (z ion)	-	**188.0706**	**165.0546**	-
Others	-	**170.0600**; **146.0600**	-	-

Table S3. Product ions deriving from modified amino acid fragmentations. Fragments in green were inserted in compound class scoring tool based on experimental data or spectra reported in the literature.

	Hydroxyproline (Hyp)	**Hydroxylysine (Hyl)**	**Pyroglutamic acid (Pyr)**	**Citrullin (Cit)**	**Methylarginine (MeArg)**
Formula	$C_5H_9NO_3$	$C_6H_{14}N_2O_3$	$C_5H_7NO_3$	$C_6H_{13}N_3O_3$	$C_7H_{16}N_4O_2$
Molecular Weight	131.058244	162.100442	129.042594	175.095691	188.127326
$[M+H]^+$	132.065520	163.107718	130.049870	176.102967	189.134602
Iminium Ion (a ion)	**86.0600**	117.1022	**84.0444**	130.0975	143.1291
Im – H_2O	**68.0495**	**99.0917**	-	**112.0869**	-
Im – NH_3	-	**100.0757**	-	**113.0709**	**126.1026**
b ion	**114.0550**	**145.0972**	**112.0393**	**158.0924**	**171.1240**
b – H_2O	**96.0444**	**127.0866**	-	**140.0818**	-
b – NH_3	97.0284	**128.0706**	-	141.0659	154.0975
c ion	131.0815	162.1237	**129.0659**	175.1190	188.1506
x ion	158.0448	189.0870	-	202.0822	215.1139
y ion	**132.0655**	**163.1077**	-	176.1030	**189.1346**
y – H_2O	**114.0550**	**145.0972**	-	**158.0924**	-
y – NH_3 (z ion)	-	**146.0812**	-	**159.0764**	**172.1081**
Others	-	-	-	-	**112.0869**; **70.0561**
	Methyllysine (MeLys)	**Dimethyllysine (DiMeLys)**	**Trimethyllysine (TriMeLys)**	**Acetyllysine (AcLys)**	**Succinyllysine (SucLys)**
Formula	$C_7H_{16}N_2O_2$	$C_8H_{18}N_2O_2$	$C_9H_{21}N_2O_2^+$	$C_8H_{16}N_2O_3$	$C_{10}H_{18}N_2O_5$
Molecular Weight	160.121178	174.136828	189.159754	188.116093	246.121573
$[M+H]^+$	161.128454	175.144104	189.159754	189.123369	247.128849
Iminium Ion (a ion)	**115.1230**	**129.1386**	**143.1543**	**143.1179**	**201.1234**
Im – H_2O	-	-	-	-	-

Im – NH_3	**98.0964**	**112.1121**	**126.1277**	**126.0913**	**184.0968**
b ion	**143.1179**	**157.1335**	**171.1492**	**171.1128**	**229.1183**
b – H_2O	-	-	-	-	-
b – NH_3	**126.0913**	**140.1070**	**154.1226**	**154.0863**	**212.0917**
c ion	160.1444	174.1601	188.1757	188.1394	246.1448
x ion	187.1077	201.1234	215.1390	188.1394	246.1448
y ion	**161.1285**	**175.1441**	**189.1598**	**189.1234**	**247.1288**
y – H_2O	-	-	-	-	-
y – NH_3 (z ion)	-	-	-	-	-
Others	**130.0863**; **84.0808**	**130.0863**; **84.0808**	**130.0863**; **84.0808**	**130.0863**; **84.0808**	**130.0863**; **84.0808**

	Methionine sulfoxide (Mes)	**Tyrosine O-sulfate (sTyr)**	**Serine O-phosphate (pSer)**
Formula	$C_5H_{11}NO_3S$	$C_9H_{11}NO_6S$	$C_3H_8NO_6P$
Molecular Weight	165.045966	261.030711	185.008890
$[M+H]^+$	166.053242	262.037987	186.016166
Iminium Ion (a ion)	**120.0478**	**216.0325**	**140.0107**
Im – H_2O	-	-	-
Im – NH_3	103.0212	-	-
b ion	**148.0427**	**244.0274**	**168.0056**
b – H_2O	-	-	-
b – NH_3	131.0161	-	-
c ion	165.0692	261.0540	185.0322
x ion	192.0325	288.0173	211.9954
y ion	**166.0532**	**262.0380**	**186.0162**
y – H_2O	-	-	-
y – NH_3 (z ion)	**149.0267**	**245.0114**	168.9896
Others	**56.0495**; **84.0444**; **102.0550**	**136.0757**; **165.0546**; **182.0812**	**60.0444**; **106.0499**

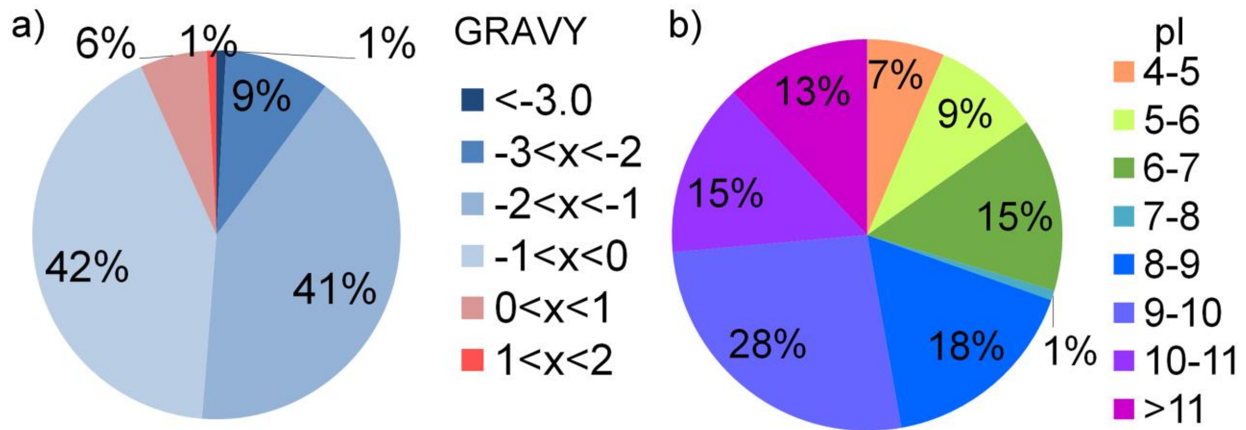


Figure S3. Physico-chemical properties of the peptides identified by pNovo and endogenous in fish muscle, describing (a) hydrophilicity calculated as GRAVY index and (b) isoelectric point.

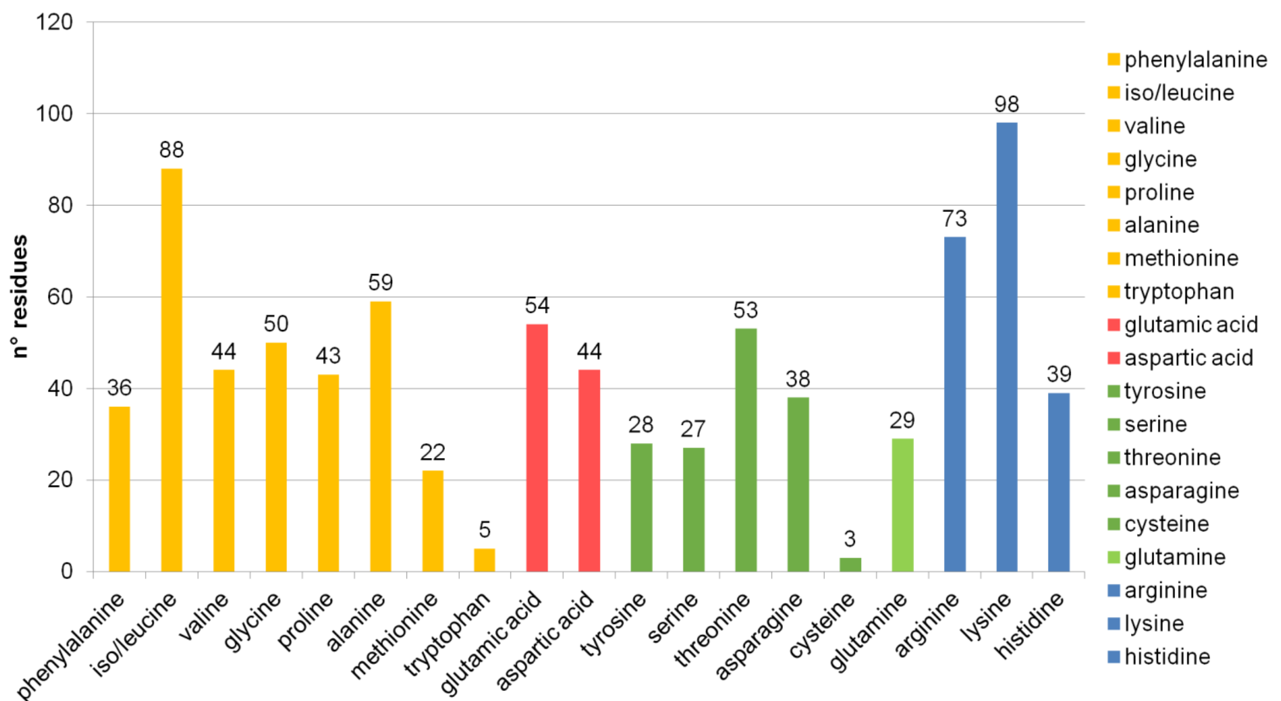


Figure S4. Occurrence of the single amino acid in the sequences of the peptides identified by pNovo and endogenous in fish muscle. Colors refer to the nature of the amino acid side chain, i.e. hydrophobic (yellow), acid (red), polar (green), basic (blue).

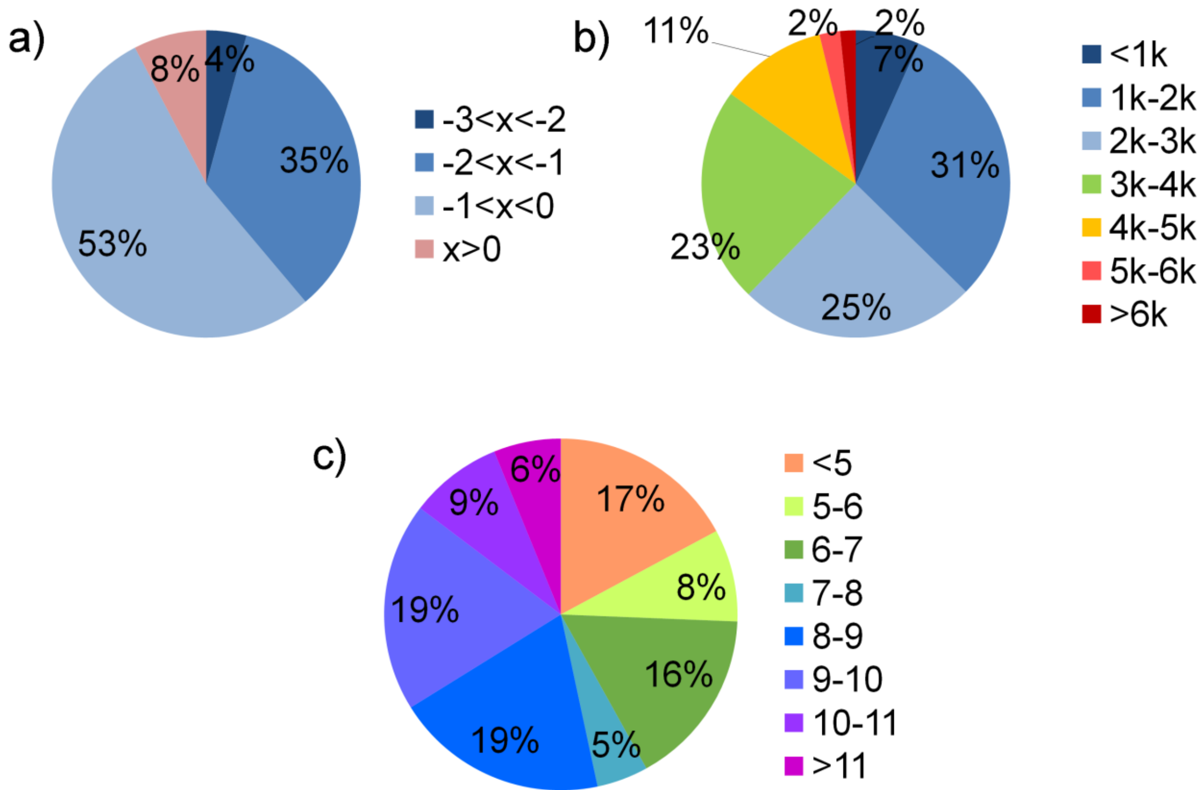


Figure S5. Physico-chemical properties of the peptides identified by PEAKS database search and endogenous in fish muscle, describing (a) hydrophilicity, calculated as GRAVY index, (b) molecular mass range (Da) and (c) isoelectric point.

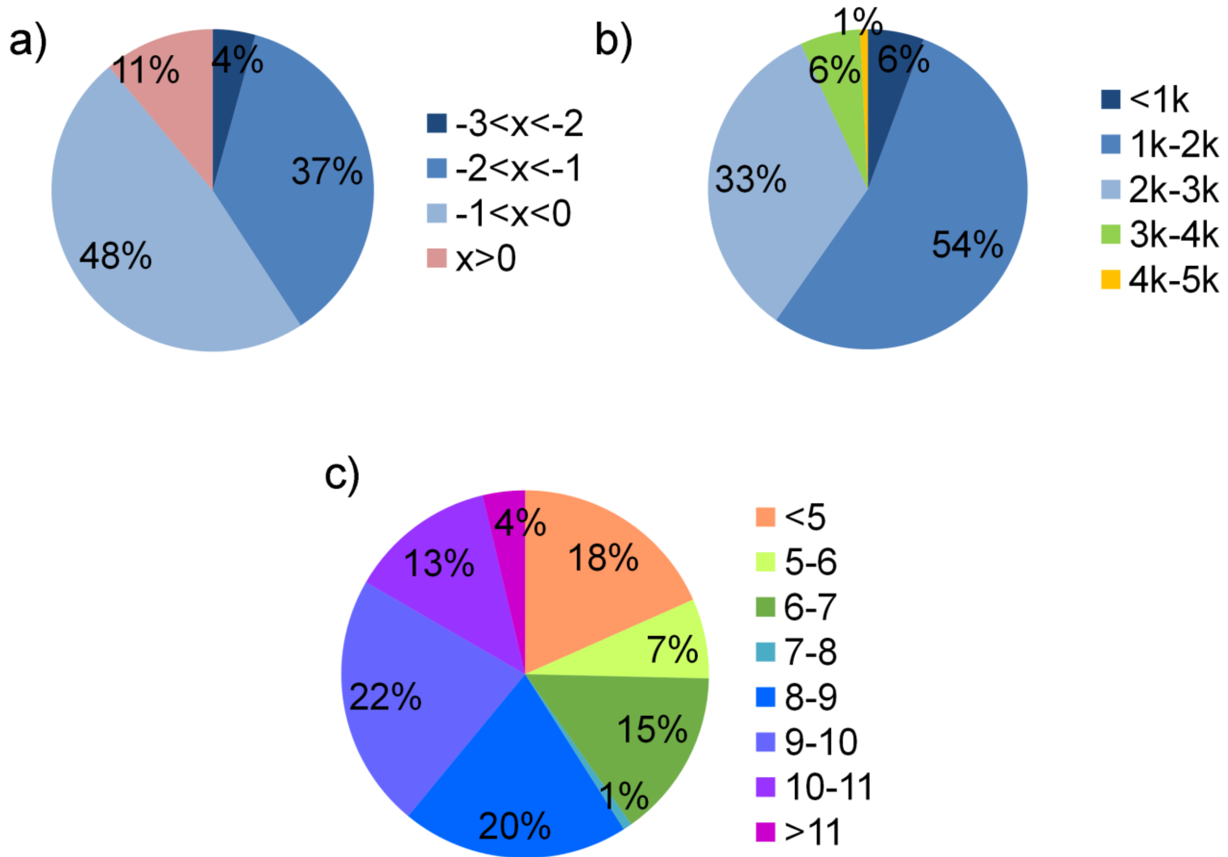


Figure S6. Physico-chemical properties of the peptides identified by PEAKS *de novo* search and endogenous in fish muscle, describing (a) hydrophilicity, calculated as GRAVY index, (b) molecular mass range (Da) and (c) isoelectric point.

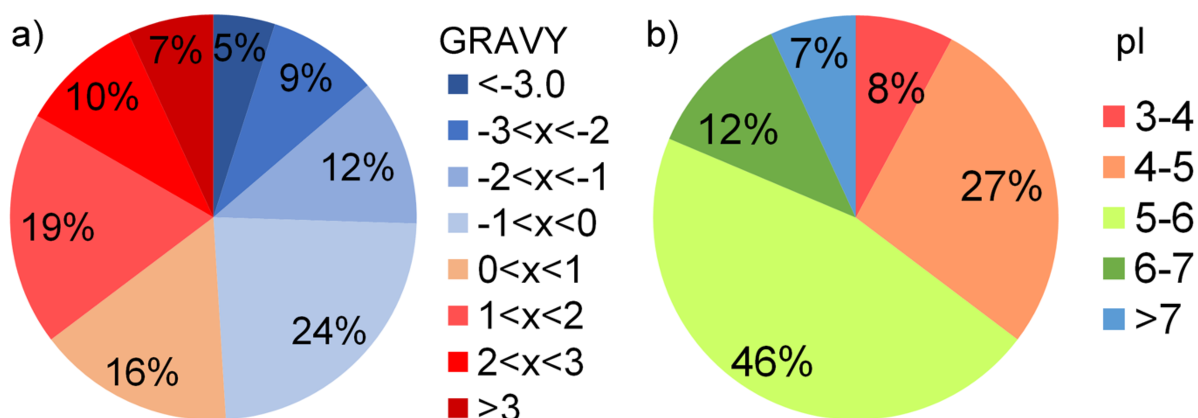


Figure S7. Physico-chemical properties of the identified short peptides endogenous in fish muscle, describing (a) hydrophilicity, calculated as GRAVY index and (b) isoelectric point.

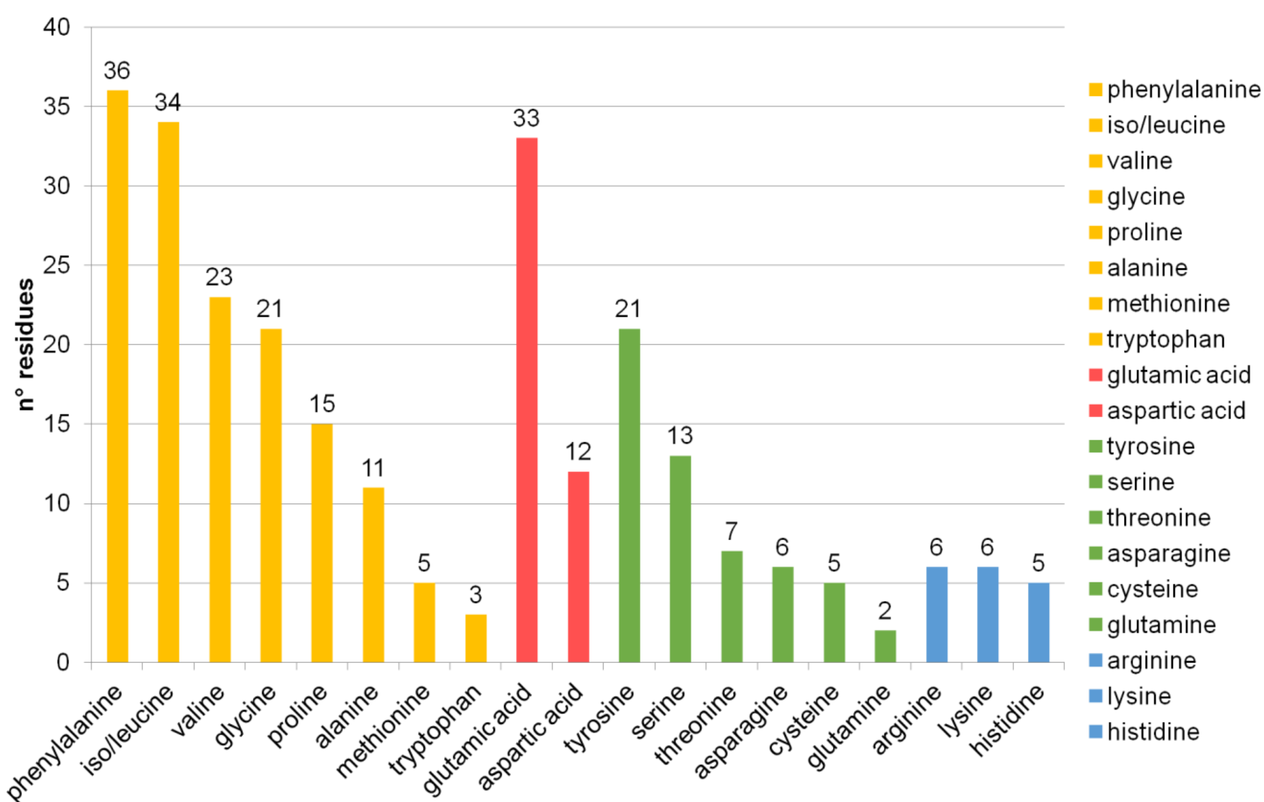


Figure S8. occurrence of the single amino acid in the sequences of the short peptides identified from fish muscle endogenous peptides. Colors refer to the nature of the amino acid side chain, i.e. hydrophobic (yellow), acid (red), polar (green), basic (blue).

Supplementary Material
Click here to download Supplementary Material: Supporting information Tables S4-13.xlsx

Table S4. Peptide identifications provided by pNovo for medium-sized peptides.
For each peptide, the following information is provided: sequence, modification, PSM score
mass , delta mass (ppm), score for each amino acid, , GRAVY (grand average of hydropathy)

Sq	Mod_Sites
AANLQSKSFR	
AAVAAQGKAKK	
AEAAMFHR	
AERFAASSR	
ALSDSETKAFLK	
ALTDAETK	
APLLAVTR	
CHNDLKMK	
CPSNLGTKR	8,Acetyl[K];
EAHLYR	0,Glu->pyro-Glu[AnyN-termE];
EAYDKSSSYTVR	
ELYGKLR	
FEELLNR	
FEELLNRLR	
FEELLTR	
FPDLSLHNNHMAK	
FVLETGR	
FVLETGRQAK	
GHADLMQQK	
HGEEQVKLWR	
HGESAWNQENR	
HGLYEKKKTSR	
HLKVVYR	
HMGYGKR	
HNDLKMK	
HNNHMAKVLTK	
KAAKPSKPK	
KAGAAEKGVPLYR	
KAHRDRKKPR	6,Methyl[R];
KAPTKETFR	
KAPVKKPK	
KAQAEKGVPLYR	
KASEFFR	1,Acetyl[K];
KDHVNKDLAPK	1,Acetyl[K];
KDQGKAKPST	1,Succinyl[K];
KEEFPNLSK	1,Succinyl[K];
KELYGKLR	
KEQEQLAALR	
KEVVPAKTEAKSK	

***Declaration of Interest Statement**

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

CRediT Author Statement

Andrea Cerrato: Investigation, Data Curation, Writing - original draft. **Sara Elsa Aita**: Investigation. **Chiara Cavaliere**: Investigation, Supervision. **Aldo Laganà**: Resources, Funding acquisition, Supervision. **Carmela Maria Montone**: Investigation, Writing - original draft. **Susy Piovesana**: Data Curation, Writing - original draft. **Riccardo Zenezini Chiozzi**: Supervision, Data Curation. **Anna Laura Capriotti**: Conceptualization, Supervision, Project administration