



**HAL**  
open science

## Identifying Short-Chain Peptides in Agrofood: Challenges and Critical Insights

Alexandre Weber, Barbara Deracinois, Benoit Cudennec, Maxime De sousa lopes moreira, Rozenn Ravallec, Thierry Grard, Christophe Flahaut

► **To cite this version:**

Alexandre Weber, Barbara Deracinois, Benoit Cudennec, Maxime De sousa lopes moreira, Rozenn Ravallec, et al.. Identifying Short-Chain Peptides in Agrofood: Challenges and Critical Insights. 2025. hal-04937346

**HAL Id: hal-04937346**

**<https://hal.science/hal-04937346v1>**

Preprint submitted on 10 Feb 2025

**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.

# Identifying Short-Chain Peptides in Agrofood: Challenges and Critical Insights.

Alexandre Weber, Barbara Deracinois, Benoit Cudennec, Maxime de Sousa Lopes Moreira, Rozenn Ravallec, Thierry Grard, Christophe Flahaut\*

Univ. Artois, UMRt 1158 BioEcoAgro, INRAe, Univ. Littoral Côte d'Opale, Univ.Lille, Univ. Picardie Jules Verne, Univ. Liège, 62300 Lens, France

\*Corresponding author: [christophe.flahaut@univ-artois.fr](mailto:christophe.flahaut@univ-artois.fr)

## Highlights

- Nearly 100% of peptides identified by peptidomics or proteomics are large or medium-chain peptides.

- Current peptidomics and proteomics methodologies are not well-suited for identifying short-chain peptides with fewer than six amino acids.

- Identifying short-chain peptides necessitates the development of specialized peptidomics techniques tailored to their detection.

- The creation of international guidelines for the identification and reporting of short-chain peptides would help standardize practices and enhance consistency in research.

## Abstract

As the global population continues to grow, optimizing the use of the protein resources produced by our planet has become essential for supporting the nutrition and health of both humans and animals. Over the years, the agro-food industry has played a key role in valorizing protein-rich by-products generated during food processing. Fermentation and enzymatic hydrolysis of these by-products are efficient, industry-friendly, and cost-effective methods for recovering peptides in the form of peptide hydrolysates. These hydrolysates contain a range of protein compounds, including large-, medium-, and short-chain peptides (SCPs), as well as amino acids, many of which are bioactive. Numerous databases now catalog food-derived bioactive peptides, with the BioPep-UWM database being one of the

33 most comprehensive, containing 5,014 bioactive peptides (as of July 2024) associated with  
34 90 distinct bioactivities. Of these, nearly 45% are SCPs, including 581 di-, 754 tri-, 385 tetra-,  
35 and 511 penta-peptides, totaling 2,231 SCPs. Despite their prevalence, the identification of  
36 SCPs remains more challenging than that of large- or medium-chain peptides, and research  
37 dedicated to SCP identification began only about twelve years ago. This review critically  
38 examines the technical challenges involved in SCP identification, offering a detailed timeline  
39 of approximately fifty key studies that have contributed to advancing the field. Specifically,  
40 the review explores enrichment and clean-up protocols, analytical methods, mass  
41 spectrometry techniques, and the bioinformatics approaches used for tandem mass  
42 spectrometry data processing in the context of SCP identification.

43

#### 44 **Keywords:**

45 Bioinformatics, liquid chromatography, mass spectrometry, peptidomics, short-chain  
46 peptides.

47

#### 48 **Abbreviations**

49 3-NBA, 3-nitrobenzyl alcohol

50 ACE, angiotensin-converting enzyme

51 ACN, acetonitrile

52 AMT, accurate mass tag

53 CCK, cholecystokinin

54 CE, capillary electrophoresis

55 CID, collision-induced dissociation

56 DDA, data-dependent acquisition

57 DIA, data-independent acquisition

58 DPP-IV, dipeptidyl-peptidase IV

59 GCB, graphitized carbon black

60 GIP, glucose-dependent insulinotropic peptide

61 GLP-1, glucagon-like peptide-1

62 HCD, higher energy collisional dissociation

63 HILIC, hydrophilic interaction liquid chromatography

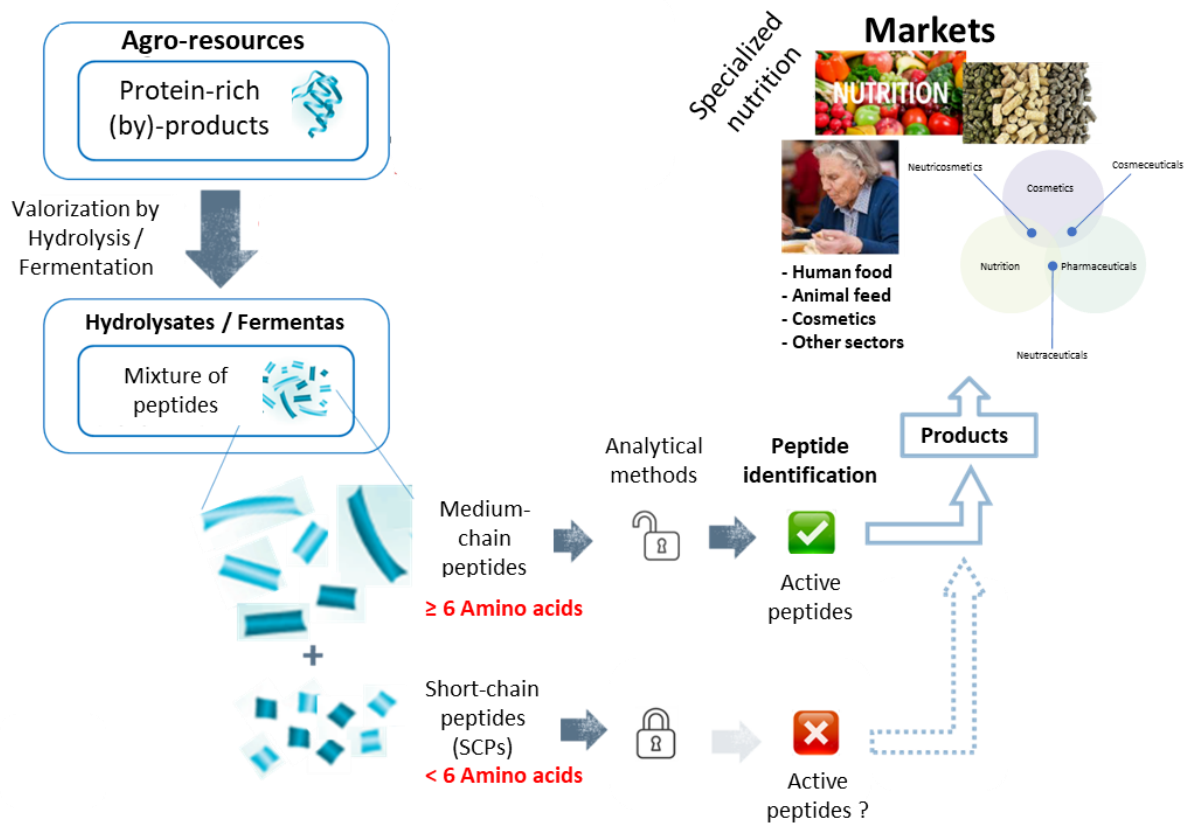
64 HPLC, high-performance liquid chromatography

65 HR, high resolution

- 66 MRM, multiple reaction monitoring
- 67 MS, mass spectrometry
- 68 MS/MS, tandem mass spectrometry
- 69 PGC, porous graphitic carbon
- 70 RF, radio frequency
- 71 RP, reverse phase
- 72 SCPs, short chain peptides
- 73 SPE, solid-phase extraction
- 74 TFA, trifluoroacetic acid
- 75 UPLC, ultra-performance liquid chromatography
- 76

77 1. Introduction

78 The sustainable consumption of land and sea resources has become a critical global  
 79 imperative. While some of these resources took millennia to accumulate, others—such as  
 80 those within the agri-food industries—naturally align with the cyclical processes of life. The  
 81 agri-food sector, deeply attuned to the need for resource efficiency and optimization, has  
 82 been undergoing a significant transformation, driven by European policies and an increasing  
 83 awareness of environmental concerns. Over recent years, the industry has shifted focus,  
 84 embracing the most thoughtful and innovative ways to valorize agro-resources. Beyond its  
 85 primary role in food production for both humans and animals, the agri-food sector has long  
 86 recognized that its by-products present a valuable opportunity for additional economic and  
 87 environmental benefits. For instance, the extraction of bioactive peptides from these by-  
 88 products—used in food, animal feed, cosmetics, and even pharmaceuticals—has emerged  
 89 as a promising avenue for innovation and added value within the industry (Figure 1).



90  
 91 Figure 1: Schematic drawing illustrating i) the valorization of protein-rich products or protein-  
 92 rich by-products through enzymatic hydrolysis or fermentation, ii) the analytical challenge of  
 93 short-chain peptide (SCP) identification in a complex mixture of peptides of various lengths,  
 94 and iii) the potential markets of bioactive peptides.

95

## 96 2. Valorization of protein-rich by-products

97 Protein-rich by-products from industrial processes can be valorized either through the  
98 proteins themselves or by utilizing industrial techniques that render the proteins soluble,  
99 typically by converting them into peptides. In the first approach, proteins remain intact until  
100 the final product, with peptides being released *in vivo* through the digestive system or gut  
101 microbial enzymes, or a combination of both. In the second approach, proteins are typically  
102 converted into peptides through processes like fermentation or, more commonly, enzymatic  
103 hydrolysis (Aspevik et al., 2017; Boboua et al., 2024). Enzyme release is a highly complex  
104 and sometimes non-specific mechanism, involving various enzymes from food processing  
105 (e.g., Alcalase, Protamex, Flavorzyme) or from the digestive system (e.g., pepsin, trypsin,  
106 chymotrypsin, elastase).

107

108 The hydrolysis process results in an intricate mixture of free amino acids and peptides—  
109 ranging from dipeptides and tripeptides to tetrapeptides and even longer chains, with  
110 considerable structural variability. This variability stems from the lack of specificity in the  
111 cleavage actions of the proteolytic enzymes involved (Figure 2). Such complexity introduces  
112 challenges for analytical methods, as these peptides exhibit diverse physical and chemical  
113 properties, such as charge state, hydrophobicity, and length. As a result, specialized  
114 separation techniques and advanced informatics tools are required, unlike proteomics  
115 approaches that do not need to identify short-chain peptides (SCPs) to characterize a  
116 protein. Consequently, the food peptidome presents significant hurdles, including **i)** the  
117 exhaustive identification of medium-chain peptides and SCPs, **ii)** their analysis within food  
118 matrices, and **iii)** their tracking in *in vivo* experiments.

119

## 120 3. Bioactive peptides

121 Among the complex mixtures of food-derived peptides lie specific sequences with notable  
122 biological properties, offering potential applications across various fields, including human  
123 and animal nutrition, cosmetics, and healthcare (Figure 1). Bioactive peptides from food  
124 sources are typically inactive within the sequence of the parent protein but, after being  
125 released through enzymatic proteolysis, can exert a range of physiological functions (Karami  
126 & Akbari-adergani, 2019).

127

128 Research into bioactive peptides is well-established, with extensive experimental data now  
129 being collated in specialized bioactive peptide databases. However, not all bioactive peptides  
130 are yet fully documented in these resources. Currently, more than 100 bioactive peptide-  
131 related databases are freely accessible online ([https://biochemia.uwm.edu.pl/bioactive-peptide-](https://biochemia.uwm.edu.pl/bioactive-peptide-databases/)  
132 [databases/](https://biochemia.uwm.edu.pl/bioactive-peptide-databases/)), compiling biological activities in relation to the structure of bioactive peptides. In  
133 the realm of food-derived bioactive peptides, one of the most widely used databases,  
134 BioPep-UWM, lists 5,014 bioactive peptides (as of July 2024), representing 90 distinct  
135 bioactivities. Notably, anti-hypertensive activity, specifically the inhibition of angiotensin-  
136 converting enzyme (ACE), is the most common feature, referenced in 1,122 peptides (over  
137 22.0%), while antioxidative activity is seen in 857 peptides (17%). Antibacterial activity  
138 follows with 526 peptides (10.5%), and anti-type-2 diabetes activity, associated with the  
139 inhibition of dipeptidyl-peptidase-IV (DPP-IV), is referenced in 489 peptides (less than 10%).

140 Among the 5,014 referenced bioactive peptides, nearly 45% are SCPs—comprising 581  
141 dipeptides, 754 tripeptides, 385 tetrapeptides, and 511 pentapeptides, totaling 2,231 SCPs.  
142 Of these, 686 SCPs (30.7%) are listed as antihypertensive, 519 (23.25%) as antioxidative,  
143 and 352 (15.75%) as anti-diabetic (inhibitors of DPP-IV). Dietary SCPs are implicated in  
144 regulating energy homeostasis through intestinal peptide sensing, which drives the secretion  
145 of gut hormones such as cholecystikinin (CCK), glucose-dependent insulintropic peptide  
146 (GIP), and glucagon like peptide-1 (GLP-1) (Duca et al., 2021), as well as modulating  
147 intestinal glucose absorption (Dugardin et al., 2022). SCPs also influence key enzymes in  
148 carbohydrate metabolism, such as DPP-IV and  $\alpha$ -glucosidase (Fleury et al., 2022). However,  
149 the relationship between peptide structure, their biological activity, and the molecular and  
150 cellular mechanisms behind these actions remains poorly understood.

151 The identification and structural characterization of bioactive peptides have garnered  
152 significant interest over the past decade, with a growing body of research aimed at  
153 elucidating the molecular and cellular mechanisms underlying their bioactivities.  
154 Approximately 1,200 scientific papers on bioactive peptides are now published annually,  
155 compared to roughly 500 a decade ago (data extracted from Web of Science). The literature  
156 contains numerous scientifically validated examples of peptides with bioactive properties  
157 (Bouglé & Bouhallab, 2017; Karami & Akbari-adergani, 2019; Mann et al., 2019).  
158 Nevertheless, the exhaustive and accurate identification of these peptides is crucial for  
159 assessing the bioactivity of peptide mixtures, necessitating further investigation into the  
160 individual effects of each peptide identified.

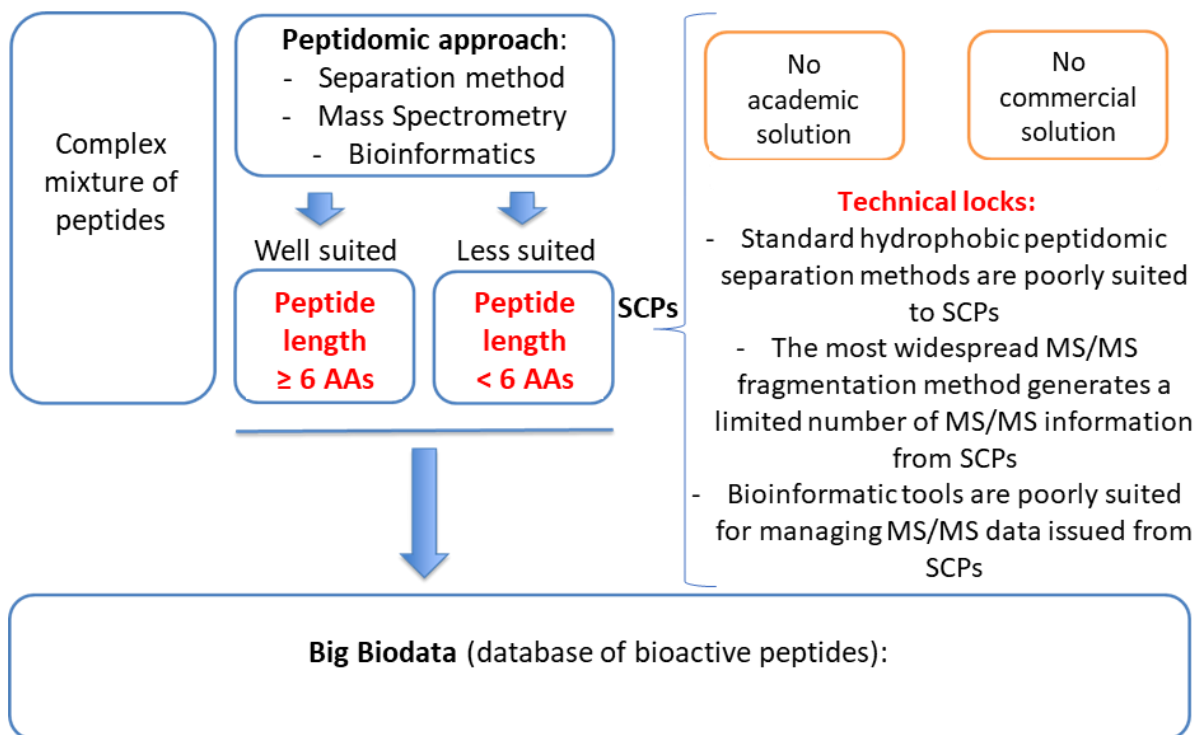
161

162 4. Peptidomics technical challenges in SCP identification

163 To drive industrial applications supported by patents, the search for bioactive peptides is  
164 closely tied to the precise structural characterization of the peptides within complex mixtures,  
165 particularly the bioactive peptides themselves. In recent years, the identification and  
166 quantification of large- and medium-chain peptides using peptidomic approaches have  
167 become more accessible, thanks to advances in **i)** (nano) high-pressure liquid  
168 chromatography (HPLC) coupled with single and tandem mass spectrometry (MS and  
169 MS/MS), **ii)** protein databases, and **iii)** bioinformatics tools. These tools provide essential  
170 connections between mass spectrometry data and protein databases, strengthening the  
171 robustness of peptide identification. However, significant limitations persist, particularly when  
172 it comes to the identification of SCPs. Indeed, current technologies and methodologies are  
173 not fully optimized for the characterization and identification of SCPs.

174 In our view, three key technological hurdles hinder the accurate identification of SCPs  
175 (Figure 2). While C18-based chromatographic separation methods (regardless of column  
176 diameter), mass spectrometry protocols (regardless of the type of MS analyzer), and  
177 bioinformatic tools for managing HPLC-run data are well-suited to the identification and  
178 quantification of large- and medium-chain peptides, they are less effective for SCP  
179 identification. Specifically, **i)** the increased hydrophilicity of most SCPs, compared to large- or  
180 medium-chain peptides, makes it difficult for conventional hydrophobic HPLC separation  
181 methods used in peptidomics to effectively separate these hydrophilic SCPs; **ii)** the most  
182 common MS/MS fragmentation techniques generates a limited amount of MS/MS data for  
183 SCPs; and **iii)** the existing bioinformatic tools are not well-suited to handle MS/MS data  
184 resulting from SCP fragmentation, leading to challenges in processing and interpreting the  
185 data accurately.





186

187 Figure 2: Schematic drawing depicting the key points of peptidomics and the technical  
 188 barriers that hinder the identification of short-chain peptides (SCPs) and amino acids (AAs).

189

190 5. Peptidomics for SCP identification

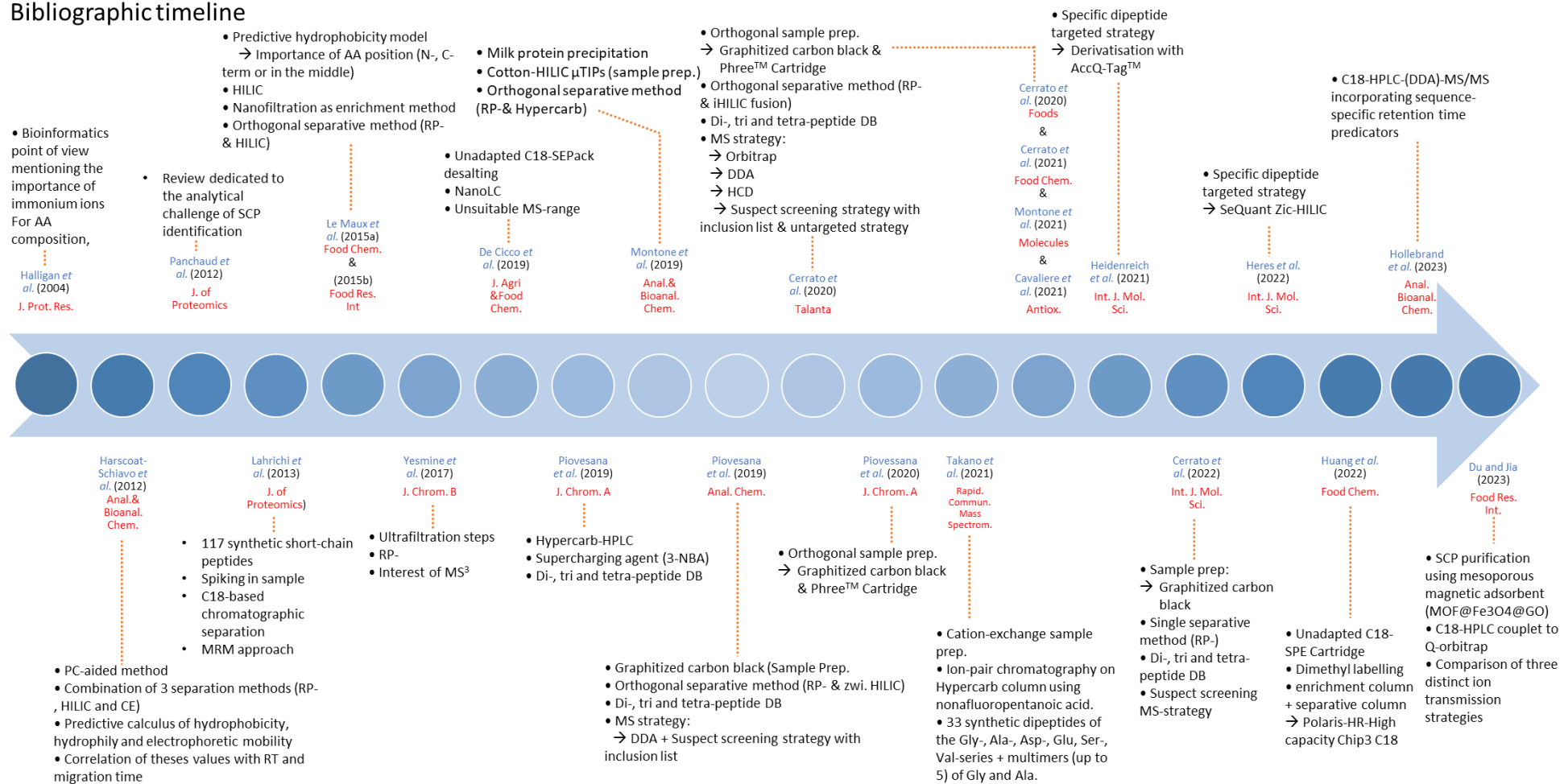
191 In this section, we provide a bibliographic timeline (Figure 3) spanning the past decade to  
 192 offer an overview of sample preparation and analytical methods reported in the literature for  
 193 SCP identification.

194 The study of SCP peptidomics began approximately twenty years ago. In 2004, Halligan *et al.*  
 195 introduced, from a bioinformatics perspective, the challenge of identifying peptides  
 196 consisting of a small number of amino acids (Halligan *et al.*, 2004). This led to the  
 197 development of innovative vector-based approaches for peptide identification by various  
 198 bioinformatics research groups (Cittaro *et al.*, 2005).

199 From a practical standpoint, the analytical task of identifying SCPs, especially di-, tri-, and  
 200 tetra-peptides, and the challenges that arose at the interface between peptide and small  
 201 molecule analysis (metabolomics), became more prominent just over a decade ago (**Erreur !  
 202 Source du renvoi introuvable.**). Panchaud *et al.* highlighted that “while SCPs containing  
 203 some hydrophobic amino acids are usually retained on classical C18 columns, SCPs  
 204 composed of only hydrophilic amino acids are not” (Panchaud *et al.*, 2012). Therefore,  
 205 separating SCPs before injecting them into the mass spectrometer requires a combination of

206 different techniques. Once separated, peptides are analyzed by mass spectrometry, but their  
207 characterization is far from straightforward. During this stage, two critical parameters  
208 emerge: **i)** high mass accuracy, and **ii)** the collision energy used for fragmentation  
209 (Panchaud et al., 2012).

## Bibliographic timeline



210

211 Figure 3: Bibliographic timeline of the last 20 years, showing analytical methods reported for short-chain peptide identification.

212

213

214 Panchaud *et al.* discussed these limitations and proposed a pragmatic yet time-consuming  
215 solution based on HPLC-MS/MS data acquisition from a library of synthetic SCPs (Panchaud  
216 *et al.*, 2012). In their study, 117 peptides predominantly composed of branched-chain amino  
217 acids, were synthesized and used in a case study. A C18-based HPLC approach coupled  
218 with a data-dependent analysis, specifically multiple reaction monitoring (MRM), was employed  
219 to separate and characterize the synthetic SCPs (Lahrichi *et al.*, 2013). This approach was  
220 also applied to characterize the SCP heterogeneity in milk hydrolysate. The authors  
221 concluded that even for a highly complex peptide set, HPLC-MS/MS utilizing a MRM  
222 strategy is an effective method for discriminating peptides. This is achieved through a  
223 combination of retention time, diagnostic ions, and ratios of key diagnostic ions. However,  
224 while diagnostic ions allow for the unambiguous identification of dipeptides and at least the  
225 N-terminus of tripeptides, the analysis becomes more challenging for tetrapeptides when  
226 both leucine (Leu) and isoleucine (Ile) are present.

227

228 In the same year (2012), Harscoat-Shiavo *et al.* introduced a computer-aided method  
229 designed for SCP characterization. This method combined three separation techniques—  
230 reverse phase (RP) chromatography, hydrophilic interaction liquid chromatography (HILIC),  
231 and capillary electrophoresis (CE)—with predictive calculations of hydrophobicity,  
232 hydrophilicity, and electrophoretic mobility. The authors demonstrated that these values  
233 correlated with chromatographic retention times and migration times (Harscoat-Schiavo *et al.*  
234 *et al.*, 2012). They concluded that the selectivity characteristics of these three complementary  
235 separation techniques, when used in combination, significantly reduced the possible number  
236 of peptide combinations (by as much as 98%). A major advantage of the proposed method is  
237 the physicochemical information provided by such analyses. Since biological activities are  
238 often linked to specific physicochemical properties (Pripp, 2007), these analyses could assist  
239 in the discovery of bioactive peptides. However, it is worth noting that the proof of concept for  
240 this methodology was based on a set of only 18 synthetic SCPs, and no justification was  
241 provided for the selection of these specific SCPs.

242

243 Three to four years later, Fitzgerald's research group developed a method to improve SCP  
244 identification based on a predictive model of retention time for these peptides on HILIC  
245 coupled to MS/MS (Le Maux, Nongonierma, & Fitzgerald, 2015). The reported predictive  
246 algorithmic model for chromatographic retention time was trained using the length of 153  
247 SCPs and the position (N-, C-terminal, or central) of amino acids in the peptide sequence.

248 This model was validated using 14 synthetic peptides (excluding 10% of the training SCP  
249 set) and ultimately linked the SCP retention time to the apparent hydrophilicity coefficient and  
250 the peptide sequence size. The authors provided experimental validation of the model using  
251 a solid-phase extraction (SPE) fraction of whey protein hydrolysate analyzed with an MS  
252 method for SCP detection ( $m/z$  range of 21–600), confirming the model's applicability for  
253 SCPs. However, the experimental validation was performed on SCPs purified using a  
254 hydrophobic SPE fraction, despite the unbound hydrophobic SPE fraction being reported by  
255 Nongonierma *et al.* 2013 (Nongonierma et al., 2013). The method should have been  
256 evaluated using the entire whey protein hydrolysate sample, rather than just a subfraction.  
257 Moreover, the 14 synthetic peptides used as standards could have been spiked into the  
258 whey protein hydrolysate to assess the efficiency of the SPE enrichment and the influence of  
259 the matrix effect on both the SPE fractionation and the analytical method.

260

261 Simultaneously, Fitzgerald's group applied this predictive algorithmic model of peptide  
262 retention time on HILIC in combination with a second predictive model of peptide retention  
263 time on RP chromatography, using the same set of 153 peptide sequences (Le Maux,  
264 Nongonierma, & Fitzgerald, 2015; Le Maux, Nongonierma, Murray, et al., 2015). This two-  
265 dimensional chromatographic separation approach was validated using eight free amino  
266 acids, twenty dipeptides (composed of A, V, I, G, L, W, and Y), and two tetrapeptides. The  
267 authors concluded that the elution order of peptides under RP chromatography and HILIC  
268 conditions was not symmetrical, highlighting the complementary nature of these two  
269 separation techniques. The differences observed between the two methods demonstrated  
270 that HILIC and RP-HPLC could work synergistically to enhance peptide separation. This  
271 approach was specifically optimized to differentiate between predominantly homologous  
272 dipeptides, which are often challenging to identify through MS/MS or *de novo* searches.  
273 However, from our perspective, the synthetic peptide dataset was tailored to address a  
274 specific research question—namely, identifying SCPs with antidiabetic properties by  
275 inhibiting cellular and circulating dipeptidyl peptidase IV (DPP-IV).

276 The Fitzgerald's group also published a sort communication on optimizing MS/MS  
277 fragmentation conditions for SCP characterization using micrOTOF QII-based mass  
278 spectrometry (O'keeffe & Fitzgerald, 2015). In this study, the authors employed a range of  
279 synthetic di-, tri-, and tetra-SCPs separated on C18-UPLC and analyzed by MS/MS to fine-  
280 tune the fragmentation conditions, reducing over-fragmentation that typically occurs with  
281 standard MS/MS protocols. Specifically, they narrowed the mass range to allow the transfer  
282 of precursor ions with  $m/z$  values between 50 and 600, minimized the collision cell radio

283 frequency (RF) to 140 Vpp, and adjusted the transfer time to 70  $\mu$ s. These adjustments  
284 resulted in more efficient transfer and optimal fragmentation of SCPs. However, the  
285 identification of SCPs still requires extensive manual interpretation of the MS/MS spectra,  
286 and the authors expressed hope that ongoing advancements in software will alleviate the  
287 time-consuming nature of this process (Figure 3).

288

289 In 2017, Yesmine *et al.* reported the results of an accurate mass tag (AMT) study aimed at  
290 the characterizing a subset of short angiotensin-I-converting enzyme (ACE) inhibitory  
291 peptides. This study involved the creation of a C18-UPLC-ESI-MS/MS database based on  
292 retention time, the accurate mass of parent ions, and the most intense fragment ions from 11  
293 di- and tri-peptides previously identified for their ACE-inhibitory effects (Yesmine *et al.*,  
294 2017). The primary goal of the study was to detect these 11 SCPs in a nanofiltration-based  
295 subfraction of Tilapia (*Oreochromis niloticus*) protein hydrolysate, but only two of the  
296 peptides were detected and unequivocally identified using the BioLynx module and Fragment  
297 Ion Calculator software. Additionally, the authors reported the *de novo* identification of 51  
298 other SCPs based on their collision-induced dissociation (CID) patterns. Both the peptide  
299 sequencing module of Waters MassLynx.V. 4.1 software and manual calculations were  
300 employed to process the MS/MS data and perform SCP sequencing. A key limitation of the  
301 method was the inability to distinguish between isoleucine and leucine in the amino acid  
302 sequences of the SCPs. Furthermore, the authors noted that the m/z range (50–1200) and  
303 scan time (0.5 sec) used in their analysis could have been further optimized.

304

305 Two years later, De Cicco *et al.* published the results of their study focused on characterizing  
306 the hidden digestome of whey proteins subjected to *in vitro* human gastrointestinal digestion  
307 (Yesmine *et al.*, 2017). The term “hidden digestome” or “hidden peptidome resulting from  
308 gastrointestinal digestion” aptly refers to the array of large-chain peptides, SCPs, and cross-  
309 linked peptides such as those containing disulphide bridges or desmosine/isodesmosine  
310 bridges. To enhance the identification of these “hidden” peptides, the authors utilized a multi-  
311 step approach: **i)** a protein precipitation protocol, **ii)** a nanochromatographic column, **iii)** a  
312 long elution gradient to improve chromatographic separation, **iv)** reduction of disulphide  
313 bridges followed by reanalysis of the reduced samples under the same conditions, and **v)**  
314 various peptide identification software. Overall, the peptides identified were larger than five  
315 amino acids, likely due to C18-SPE-based peptide purification prior to nanoHPLC-MS/MS  
316 analysis, with a 200–1200 m/z range used. This article also highlights that the definition of

317 SCPs may vary among scientists. In our view, SCPs should be defined as peptides  
318 containing between two and five amino acids.

319 The elimination of “contaminant” molecules from the analyte of interest is a crucial step in  
320 analytical sciences. This clean-up process minimizes undesirable molecular interferences  
321 that can hinder separation. Piovesana *et al.* (2019) developed a solid-phase extraction  
322 method to clean and enrich SCPs (di-, tri-, and tetra-peptides) from native urine, without the  
323 need for derivatization (Piovesana, Capriotti, et al., 2019). The proof of concept and method  
324 efficacy were evaluated using a set of 13 synthetic di- to tetra-peptides (6 di-, 4 tri-, and 3  
325 tetra-peptides). The SCPs were isolated using graphitized carbon black (GCB) Carbograph 4  
326 sorbent, requiring fine-tuning of the short peptide elution process, which resulted in recovery  
327 rates ranging from 60 to 100%. However, in our opinion, the choice of amino acid sequences  
328 for the peptides in this study was not fully explained, the enrichment in SCPs was not  
329 verified, as no quantitative comparison between spiked SCPs and all urinary molecules  
330 (including larger peptides) was performed, and the 13 synthetic peptides used represent a  
331 small sample compared to the theoretical 168,400 combinations for di-, tri-, and tetra-  
332 peptides (400 dipeptides, 8,000 tripeptides, and 160,000 tetrapeptides). Nonetheless, this  
333 study marked the first time that pre-purified SCPs were analyzed in triplicates using an  
334 orthogonal UHPLC-high resolution (HR) MS/MS approach, employing both RP C18 and  
335 zwitterionic HILIC columns. The SCPs were identified based on HR-MS/MS data obtained  
336 with a hybrid quadrupole-Orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific),  
337 with data-dependent acquisition (DDA) mode, suspect screening, inclusion lists, and higher-  
338 energy collisional dissociation (HCD). Peptides were identified using a semi-automated  
339 workflow for feature extraction, candidate mass filtering, and MS/MS spectra comparison  
340 with *in silico* mass spectra. Based on the largely unknown number of SCPs in urine, the  
341 authors reported identifying 152 peptides (101 from RP runs, 111 from HILIC runs, with 60  
342 common identifications). The same team later expanded their SCP database by  
343 incorporating 13 modified amino acids in addition to the 20 standard proteinogenic ones.  
344 They also demonstrated that using an MS inclusion list combined with suspect screening in  
345 DDA mode was much more effective for identifying SCPs than the data-independent  
346 acquisition (DIA) mode, leading to the tentative identification of over 200 natural and modified  
347 SCPs.

348

349 A supercharging agent, such as 3-nitrobenzyl alcohol (3-NBA), was employed in the post-  
350 column phase of the chromatographic method to address the issue of limited ionization of  
351 SCPs and mitigate ionization suppression caused by the use of trifluoroacetic acid (TFA) in

352 the mobile phases. TFA was preferred over formic acid due to its superior ion-pairing  
353 capability, which is essential for the separation of small ionizable compounds. To enhance  
354 electrospray ionization without compromising chromatographic separation, 3-NBA was  
355 added, resulting in a consistent increase in signal intensity for all analytes, up to 18-fold  
356 (Piovesana, Montone, et al., 2019).

357

358 In a parallel study, Montone *et al.* (2019), part of the same research group as Piovesana *et*  
359 *al.*, proposed a purification step for SCP characterization in cow milk to avoid suppression  
360 phenomena caused by contaminants like carbohydrates or ions. After defatting the milk  
361 through centrifugation and precipitating proteins using organic solvents, the authors  
362 employed a cotton-HILIC microtip-based purification step to enrich the sample in SCPs  
363 (Montone et al., 2019). As with Piovesana *et al.* (2019), the team utilized an orthogonal  
364 chromatographic approach, leveraging the distinct selectivity mechanisms of two different  
365 chromatographic phases: a RP (C18) phase and a porous graphitic carbon (PGC) phase.  
366 SCPs were identified using HR-MS/MS on a Q Exactive mass spectrometer, with suspect  
367 screening conducted in top-five DDA mode, in the 150–2000 m/z range. The mass data were  
368 processed using a database containing all possible di-, tri-, and tetrapeptides to identify  
369 precursor ions. Experimental HCD fragmentation spectra were matched with *in silico*  
370 fragmentation spectra, focusing on a, b and y ions. Special attention was given to iminium  
371 ions, as they are indicative of certain amino acids. Manual validation of amino acid  
372 sequences was performed based on the total number of matched ions, with a focus on the  
373 most intense ones. Validation of the orthogonal chromatographic method was conducted  
374 using a set of 14 synthetic di- to tetra-peptides (including the same 13 peptides from  
375 Piovesana *et al.*, 2019, plus one additional peptide, RKKH). The authors reported that all 14  
376 synthetic peptides were successfully separated and eluted with well-resolved peaks. The  
377 results confirmed that the purification step effectively increased the number of peptide  
378 identifications while preventing ion suppression from the presence of carbohydrates. Through  
379 this method, a comprehensive peptidomic profile of cow milk SCPs was obtained, identifying  
380 72 SCPs in total.

381

382 A comprehensive study on the extraction and clean-up of circulating endogenous SCPs in  
383 plasma was conducted by Piovesana *et al.* in 2020, where four different analytical clean-up  
384 methods were tested and compared to maximize SCP identification (Piovesana et al., 2020).  
385 The most effective approach combined protein precipitation with phospholipid removal using  
386 the Phree™ Phospholipid Removal Cartridge, followed by SCP clean-up using GCB



387 Carbograph 4. The same set of 13 synthetic peptides used in their previous studies was also  
388 referenced, though its purpose remained unclear. Chromatographic separation of both  
389 hydrophilic and hydrophobic SCPs was performed using UPLC coupled to zwitterionic HILIC,  
390 followed by HR-MS/MS analysis. This untargeted investigation was based on suspect  
391 screening in top-five DDA mode with HCD fragmentation, spanning the range of 150–1000  
392 m/z. The SCP identification process mirrored the approach from previous work. The authors  
393 reported identifying 89 SCPs, including 22 tripeptides and 57 dipeptides, using the optimized  
394 combination of purification steps. However, the identification of just one, six, and seven  
395 additional unique peptides using the other three extraction and clean-up strategies was also  
396 noted. Notably, the study did not involve synthetic peptide spiking in plasma to demonstrate  
397 the method's efficiency.

398

399 In the same year, the same research team published another study focusing on the  
400 identification of SCPs from a digestome created by the *in vitro* gastrointestinal digestion of  
401 Yellowfin Tuna (*Thunnus albacares*) fillet proteins. The proteins were extracted using liquid  
402 nitrogen crushing, cold lysis buffer, centrifugation, and trichloroacetic acid-based precipitation  
403 (Cerrato, Capriotti, et al., 2020). The SCP digestome was extracted and purified with GCB  
404 sorbent using the Carbograph 4 cartridge and chromatographed on a C18 phase, with TFA as  
405 the ion-pairing agent. The MS and MS/MS data were collected in positive ionization mode,  
406 ranging from 150–750 m/z, with HCD-based MS/MS spectra acquired using a suspect  
407 screening strategy in top-five DDA mode. An inclusion list of 4,980 unique singly charged  
408 precursor ions was used to target specific m/z values. The identification of SCPs followed the  
409 same process as outlined in their previous work (Cerrato, Aita, et al., 2020). Despite the *in*  
410 *vitro* gastrointestinal digestion of the proteins, the authors reported a substantial number  
411 (572) of “tentative” SCP identifications, as they described, indicating the high efficiency of the  
412 analytical workflow. However, the decision to use only a C18 column for HPLC separation of  
413 SCPs remains unclear, particularly when considering the team's prior successes with other  
414 chromatographic methods.

415

416 The same research team conducted an investigation into the endogenous peptides produced  
417 by peptidases after the death of sea bass (*Dicentrarchus labrax*) fillet proteins, using the  
418 same analytical workflow (Cerrato et al., 2021). A novel aspect of this study was the manual  
419 filtering of peptides based on a peptide spectrum match (PSM) score threshold of 99, which  
420 corresponds to a 1% false discovery rate (FDR). Each MS/MS spectrum was then manually  
421 checked using the graphical interface of the freeware pNovo, compared to the use of

422 Compound Discoverer software for *de novo* peptide sequencing and SCP identification. The  
423 authors concluded that while *de novo* peptide sequencing software can aid in SCP  
424 characterization, its current utility is somewhat limited unless such software is specifically  
425 adapted to the needs of SCP sequences. Ultimately, 102 unmodified SCPs (39 di-, 45 tri-,  
426 and 18 tetra-peptides) and 12 post-translationally modified peptides were identified, many of  
427 which are associated with multiple reported bioactivities.

428

429 Using the same analytical methodology, the authors also studied the endogenous short-  
430 chain peptidome profile of Italian San Daniele ham, a dry-cured ham, at three different  
431 ripening stages (14, 22, and 34 months) (Cerrato et al., 2022). The aim of this study was to  
432 statistically compare the SCP populations at different stage of ripening and identify the SCPs  
433 that influence the sensory and functional properties of dry-cured ham. After homogenizing  
434 the dry-cured ham samples, buffer-based extraction with agitation and sonication,  
435 centrifugation, and ACN-based precipitation of proteins, lipid removal was achieved using  
436 hexane-based liquid-liquid extraction. The aqueous phase was then acidified with TFA to pH  
437 2, and SCPs were purified onto Carboglyph 4 cartridges. The identification of the purified  
438 SCPs was performed using the same methodology as in the earlier studies.

439 A total of 236 SCPs were identified, including 173 natural peptides (comprising 80 di-, 48 tri-,  
440 and 45 tetra-peptides), along with 63 peptides containing non-proteinogenic amino acids,  
441 which were tentatively identified in dry-cured ham. In a separate study, but with a different  
442 focus, the quantification of seven specific dipeptides (PA, GA, VG, EE, ES, DA, and DG)  
443 demonstrated the advantages of using HILIC for salt removal as a crucial clean-up step  
444 before dipeptide ionization in the ion source of triple quadrupole mass spectrometry (Heres  
445 et al., 2022). Peptides were extracted from biceps femoris muscles excised from the ham  
446 pieces using 0.01% HCl, deproteinized via overnight precipitation with ethanol following a  
447 method published in 2015, and fractionated through ultrafiltration using Amicon® Ultra (Merk  
448 Millipore Ltd., Cork, Ireland) with a 3 kDa cutoff. The resulting permeate underwent HPLC-  
449 MS/MS analysis on a SeQuant ZIC®–HILIC capillary column (5 µm, 150 mm x 0.3 mm; Merk  
450 KGaA, Darmstadt, Germany) coupled to an ESI-QQQ 6420 Triple Quad HPLC/MS (Agilent).  
451 Elution solvents included 10 mM ammonium acetate and ACN. The highest concentration  
452 was found for GA (44.88 µg.g<sup>-1</sup> dry-cured ham), while PA had the lowest concentration (0.18  
453 µg.g<sup>-1</sup> dry-cured ham), with the remaining dipeptides ranging between 2 and 8 µg.g<sup>-1</sup> of dry-  
454 cured ham.

455

456 In a completely different field, Takano *et al.* (Takano *et al.*, 2021) reported a seamless  
457 analytical method for SCP identification, applying cation-exchange and ion-pair  
458 chromatography coupled with HR-MS/MS analysis in the geochemistry domain. The study  
459 focused on 33 synthetic dipeptides, including six starting with glycine (GG, GD, GE, GA, GS,  
460 and GV), six with alanine (AA, AG, AA, AD, AE, and AV), five with aspartate (DS, DG, DA,  
461 DD, and DV), six with glutamate (ES, EG, EA, ED, EE, and EV), five with serine (SS, SA,  
462 SD, SE, and SV), five with valine (VS, VA, VG, VD, and VE), along with three glycine-mers  
463 (GGG, GGGG, and GGGGG) and the tripeptide AAA. These synthetic SCPs were subjected  
464 to ion-pair liquid chromatography coupled with an electrospray ionisation triple Quadrupole  
465 MS system. The analysis used a reversed-phase Hypercarb stationary phase (4.6 x 150 mm,  
466 particle size of 5  $\mu\text{m}$ , porous graphitic carbon) and guard columns. Solvent A consisted of 20  
467 mM nonafluoropentanoic acid in pure water with 0.1% formic acid, while solvent B was  
468 acetonitrile with 0.1% formic acid. Synthetic SCPs were eluted with a linear gradient of 1%  
469 solvent B per minute over 60 minutes, followed by a column flush with 60% solvent B for 60  
470 minutes, and a 30-minute re-equilibration. The authors reported that all the synthetic SCPs  
471 were perfectly eluted between 26.7 and 40.7 minutes, exhibiting a narrow Gaussian peak  
472 shape. They also observed a linear retention time for the glycine series from dimer to  
473 pentamer. Given the potential interference from organic and inorganic substances in  
474 calcareous, siliceous, microbial, and sedimentary samples, the authors developed a  
475 procedure to overcome the sample matrix effect that causes ion suppression in HPLC-MS  
476 devices (Furey *et al.*, 2013). This was achieved through an isolation and desalting procedure  
477 for SCPs based on cation-exchange chromatography, using an analytical-grade AG-50 W-X8  
478 (200-400 mesh) cation-exchange resin (Takano *et al.*, 2010). The recovery of the dipeptide  
479 quantification was higher than  $94.0 \pm 4.9\%$  at a concentration of 4 ppm for each dipeptide ( $n$   
480 =10), with the mixed solution containing glyceryl series including neutral (-Gly, -Leu and -Ser),  
481 acidic (-Asp and -Glu), basic (-Arg), sulfur-containing (-Met), aromatic (-Phe), and  
482 heterocyclic (-Pro, -His, and -Trp) amino acids.

483  
484 The goal of Italian team's study was the comprehensive characterization of SCPs and  
485 medium-chain peptides derived from the simulated *in vitro* gastrointestinal digestion of soy  
486 flour protein extraction (Cavaliere *et al.*, 2021). To simplify the starting hydrolysate, the  
487 digestome was first fractionated by preparative RP-HPLC using an Xbridge BEH C18 OBD  
488 Prep (Waters) and a gradient of methanol/water/ammonium formate (pH 10). Six fractions  
489 were collected every five minutes, with fraction #2 (retention time between 5 to 10 minutes)  
490 and fraction #3 (retention time between 10 to 15 minutes) showing the most intense  
491 antioxidant activity. These fractions were then subjected to the same SCP identification  
492 methodology described earlier, resulting in the identification of 132 SCPs. However, we find it

493 somewhat illogical to fractionate the digestome using preparative RP-HPLC in the search for  
494 SCPs with antioxidant bioactivity, especially since the fractions analyzed correspond to  
495 peptides retained between five and 15 minutes on the RP(C18)-HPLC phase.

496

497 In parallel, the same Italian research team conducted an in-depth investigation of  
498 endogenous SCPs in milk and its by-products (whole milk, whole whey, skimmed whey, and  
499 whey permeate). This study utilized the previously described methodology, which included  
500 pre-enrichment of SCPs using a GCB Carbograph 4, followed by RP-HPLC-HR-MS/MS  
501 analysis and MS/MS data reprocessing for SCP identification. In total, 79 SCPs (13 di-, 31  
502 tri-, and 35 tetra-peptides) were tentatively identified after careful manual validation of the  
503 features.

504

505 For these four studies (Cerrato et al., 2021; Cerrato, Capriotti, et al., 2020), the orthogonal  
506 strategy developed a few years earlier by the same research team (Piovesana, Capriotti, et  
507 al., 2019) might have been a more suitable approach.

508

509 In addition to untargeted approaches, a dipeptide-targeted strategy has been reported as a  
510 sensitive, rapid, and reliable method for the detection and quantification of 36 dipeptides  
511 (Heidenreich et al., 2021). In this study, amino acids and dipeptides were extracted from  
512 pulverized frozen tissue using an ice-cold HCl solution (unreferenced). The samples were  
513 then labeled with the AccQ-Tag™ reagent, separated on an ACQUITY HSS T3 column with  
514 acetonitrile as the elution solvent and formic acid as a modifier. Detection and quantification  
515 were performed using the MRM mode on a QTRAP 6500+ mass spectrometer (Sciex) with  
516 the fragment mass 171.1, derived from the derivatization reagent AccQ-Tag™.  
517 Chromatographic conditions were optimized; however, more than one-third of the standard  
518 dipeptides exhibited low signal intensity, regardless of the identical concentration used for  
519 optimization (dipeptides containing multiple amine groups generally showed lower  
520 intensities). Recovery rates were evaluated between 80-100% by spiking lung tissue  
521 samples (which contain few dipeptides) with a standard dipeptide mixture at the highest  
522 concentrations tested. This approach was applied to 13 different organs, blood, and urine of  
523 wild-type mice on standard diets to reflect physiological conditions. Over the 36 dipeptides,  
524 30 were detected and quantified, while the remaining six were consistently below the  
525 detection limit. This multi-dipeptide quantification method revealed organ-specific dipeptide  
526 distribution.

528 The increasing interest in SCP identification is evident, as demonstrated by Huang *et al.*,  
529 who published a case study on almond SCPs in the same year. Their work outlines a  
530 comprehensive workflow for discovering small bioactive peptides in foods using HPLC-  
531 MS/MS (Huang *et al.*, 2022). The study begins with the extraction of protein-rich material  
532 from almond flour, employing **i**) aqueous extraction with or without Neutral Protease (*Bacillus*  
533 *subtilis*, BIO-CAT, Virginia, NY), **ii**) ethanol-based protein precipitation, and **iii**) centrifugation.  
534 Peptide dimethyl labeling was performed on both peptide standards (VY and VYV) and  
535 protein-rich extracts subjected to precipitation. The labeled and non-derivatized peptides  
536 were then purified and desalted using a C18-SPE cartridge before being introduced into an  
537 enrichment column (Agilent Polaris- HR-high-capacity Chip 3C18). Chromatographic  
538 separation was carried out with acetonitrile as the solvent and formic acid as the  
539 acidifier. MS and MS/MS measurements were conducted using an Agilent 6520 Accurate-  
540 Mass Q-TOF-MS in the 130-1400 m/z range. However, the study does not specify the data  
541 acquisition or fragmentation modes. SCP identification was performed using Peaks Studio®  
542 X+ *de novo* sequencing, with a threshold *de novo* score of  $\geq 20$ . A total of 843 SCPs were  
543 manually inspected for accurate identification. For dimethyl-labeled peptides, dimethylation  
544 was set as a fixed modification in PEAKS, except for “dimethyl labeling on Lys at any  
545 position.” Despite the methodological rigor, several concerns arise. First, while the study  
546 proposes the use of apparent a1 ions for N-terminal amino acid identification and full-length  
547 SCP sequencing, the proof-of-concept is based solely on two short-chain peptide  
548 standards—an insufficient validation in our opinion. Second, the authors do not demonstrate  
549 whether the dimethylation efficiency approaches 100%, raising questions about the suitability  
550 of C18-SPE for desalting and purification. Third, the lack of information regarding the  
551 fragmentation mode and conditions introduces uncertainty into the MS/MS data  
552 interpretation. Fourth, the *de novo* score threshold of  $\geq 20$  is notably low, casting doubt on the  
553 reliability of peptide identification. Lastly, manually verifying 843 SCPs is an arduous, time-  
554 consuming, and error-prone process. It is also regrettable that the authors did not upload  
555 their MS/MS data to ProteomExchange (<https://www.proteomexchange.org/>) and instead  
556 provided only peptide tables as supplemental material, limiting the reproducibility and  
557 transparency of their findings.

558 In early 2023, Hollebrands *et al.* introduced an improved C18-HPLC-(DDA)-MS/MS method  
559 for identifying predominantly hydrophobic SCPs, incorporating sequence-specific retention  
560 time predictors (Hollebrands *et al.*, 2023). Their approach began with swapped-sequence  
561 chemical synthesis strategy to generate a mix of 75 unique tripeptides and four sets of  
562 tetrapeptide mixture, totaling 216 distinct tetrapeptides. These five peptide mixtures were

563 dissolved in an ACN/H<sub>2</sub>O/TFA (10/89/1, v/v/v) solution at a concentration of 10 mg per 54  
564 peptides per mL (approximately 370  $\mu$ M total) and, following dilution, injected onto an Xselect  
565 peptide CSH C18 column coupled to an Ultimate 3000 RS chromatography system.  
566 Simultaneously, three food proteins (chicken egg white albumin, bovine  $\alpha$ -lactalbumin, and  $\kappa$ -  
567 casein) were hydrolyzed overnight at 37°C using papain (*Carica papaya*), and the resulting  
568 peptides were analyzed using the same instrumentation. Peptide elution was performed  
569 using a multi-step linear gradient of ACN/H<sub>2</sub>O with 0.1% FA over a total run time of 50  
570 minutes. Tri- and tetra-peptide ions were detected using a Q-Exactive Plus Hybrid  
571 Quadrupole-Orbitrap mass spectrometer in DDA mode; however, the study does not specify  
572 the fragmentation principle employed.

573 To enhance retention time prediction, the authors developed a support vector regression  
574 model incorporating chromatographic retention times, MS and MS/MS-data (peptide  
575 sequences), and two categories of molecular descriptors: **i)** non-sequence-specific  
576 descriptors (molecular mass, peptide length, theoretical average hydrophobicity (Abraham-  
577 Leo scale), isoelectric point, and eight principal component vectors derived from  
578 hydrophobic, steric, and electronic properties (VHSE scales) of 713 experimentally identified  
579 peptides (269 tri-, 360 tetra-, and 84 penta-peptides); **ii)** sequence-specific descriptors: Four  
580 indices derived from the 514 amino acid indices available in the amino acid index database  
581 (<https://www.genome.jp/aaindex/>). These included (i) the amino acid at the N-terminus, (ii)  
582 the amino acid at the C-terminus, (iii) the average of derived indices for residues between the  
583 termini, and (iv) the overall average of all amino acids in the sequence. The study concluded  
584 that incorporating sequence-specific descriptors greatly improved retention time predictions,  
585 with 93% of hydrophobic SCPs eluting in the predicted chromatographic order. However,  
586 retention times for SCPs containing additional charged side groups proved more challenging  
587 to predict.

588 A more recent study, published in May 2023, introduced a novel approach for SCP  
589 purification using a mesoporous magnetic adsorbent, MOF@Fe<sub>3</sub>O<sub>4</sub>@GO, which enables the  
590 enrichment of SCPs through both metal ion affinity and hydrophobic interactions (Du & Jia,  
591 2023). The enriched SCP extracts were separated using a Hypersil GOLD C18 column and  
592 analyzed by Q-Orbitrap (Thermo Fisher Scientific) with a mobile phase of water and ACN  
593 acidified with 0.1% TFA. To optimize SCP detection, the authors compared three distinct ion  
594 transmission strategies within the scan range of m/z 150-750: **i)** full scan plus data-  
595 dependent MS<sub>2</sub> (dd-MS<sub>2</sub>) fragmentation with an inclusion list, **ii)** full scan plus dd-MS<sub>2</sub> with  
596 both inclusion and exclusion lists, and **iii)** full scan plus dd-MS<sub>2</sub> with inclusion and exclusion  
597 lists, followed by data-independent acquisition (DIA), using a fixed isolation window of m/z  
598 104.0 with a 2 m/z overlap at the intersection of adjacent windows. MS raw data were

599 processed using Compound Discoverer, retaining only m/z values common to three  
600 replicates (excluding blank m/z) and matching ( $\pm 5$  ppm) a theoretical inclusion list of di-, tri-,  
601 and tetrapeptides. The MS/MS spectra of filtered candidate m/z values were scored based  
602 on ion fragment matching ( $\pm 5$  ppm), followed by manual validation using mMass. Among the  
603 tested strategies, the third approach provided the highest SCP screening coverage.  
604 Ultimately, 127 SCPs were identified, comprising 47 di-, 59 tri-, and 21 tetrapeptides.

605

## 606 6. Conclusions

607

608 In modern proteomics or peptidomics, the identification of large- and medium-chain peptides  
609 has become widely accessible and is routinely performed in laboratories worldwide. This is  
610 largely due to the global harmonization of experimental workflows, supported by  
611 standardized guidelines that define: **i)** pre-analytical protocols to ensure reproducibility, **ii)**  
612 optimized chromatographic methods for efficient peptide separation, **iii)** mass spectrometry  
613 operating conditions for high-resolution detection, and **iv)** bioinformatics settings integrated  
614 with protein databases for accurate peptide identification and quantification. Thanks to  
615 exponential advances in chromatographic separation techniques, mass spectrometry  
616 technologies, and bioinformatics, modern peptidomic analyses can now identify between  
617 6,000 and 10,000 peptides in a single run. However, the vast majority of these peptides are  
618 longer than 5 amino acids. This raises an important question: while peptidomics offers  
619 remarkable peptide identification capacity, its ability to detect peptides shorter than five  
620 amino acids remains significantly limited. In fact, given the current state of the field, it is  
621 estimated that over 99% of peptides identified worldwide each year belong to the large- and  
622 medium-chain categories. Given that amino acid identification techniques are well  
623 established, one pressing question remains—how can we improve the identification of short-  
624 chain peptides?

625 As reviewed here, the search for optimal analytical methods for identifying SCPs is relatively  
626 recent, particularly in the agri-food sector, emerging just over a decade ago. A key distinction  
627 from protein databases is that SCP databases can be theoretically constructed. In fact, the  
628 theoretical number of proteinogenic SCPs increases exponentially with peptide length:  
629 dipeptides 400 ( $20 \times 20$ ), tripeptides 8,000, tetrapeptides 160,000, and pentapeptides  
630 3,200,000. Despite the decreasing cost of peptide chemical synthesis, synthesizing all  
631 distinct SCPs as standards is not feasible. However, synthesizing a representative subset  
632 remains a viable approach. This raises a crucial question: why have studies in this field—  
633 except perhaps (Lahrachi et al., 2013) —relied on such a limited number synthetic peptides,

634 given the vast theoretical diversity of SCPs? A small set of synthetic peptides cannot  
635 statistically represent all possible SCP structures, yet selection criteria for these sets are  
636 rarely disclosed. Ideally, the selection of synthetic SCPs should be based on their  
637 physicochemical properties to ensure broad representation. While defining such a  
638 representative set for dipeptides is feasible, the complexity increases significantly for tri-,  
639 tetra-, and pentapeptides due to their sheer number. Moreover, considering matrix effects in  
640 real samples, any chosen synthetic SCPs should be spiked into samples to rigorously  
641 evaluate the efficiency of the analytical workflow.

642 Regardless of the sample type, an enrichment and clean-up step is essential for SCP  
643 analysis due to: i) the presence of 'contaminating' molecules (ions, saccharides, lipids, etc.)  
644 which interfere with liquid chromatography separation, regardless of the chromatographic  
645 phase used, and ii) the matrix effect, which suppresses ionization in mass spectrometer  
646 sources. Given the highly hydrophilic nature of some SCPs, C18-SPE is not recommended.  
647 Instead, GBC Carbograph 4 sorbent has been evaluated for SCP extraction from urine,  
648 achieving recovery rates of 60-100% for 13 synthetic SCPs. This sorbent was also applied in  
649 digestomic studies using *in vitro* gastrointestinal digestion of fish fillet proteins. For SCP  
650 isolation and desalting from calcareous, siliceous, microbial, and sedimentary samples, a  
651 cation-exchange chromatography approach using AG-50 W-X8 (200-400 mesh) cation-  
652 exchange resin has proven effective. In plasma, among four different clean-up methods  
653 tested for extracting endogenous circulating SCPs, the most effective approach involved: i)  
654 simultaneous protein precipitation and phospholipid removal using the Phree™ Phospholipid  
655 removal cartridge, ii) subsequent clean-up using GCB Carbograph 4. Ultimately, best  
656 strategy for enriching and purifying SCPs from peptide hydrolysates derived from food  
657 proteins is a parallel combination of at least two orthogonal chromatographic phases.

658 Initially, SCP separation by liquid chromatography relied on a single chromatographic  
659 stationary phase. However, the parallel use of two complementary phases was soon  
660 established, with one phase dedicated to the separation of hydrophilic SCPs, such as HILIC  
661 and its derivatives, and another phase for hydrophobic SCPs, such as C18 columns. In  
662 addition to these conventional approaches, porous graphitic carbon (PGC) chromatography  
663 has proven effective as a non-specific stationary phase, as described by (Piovesana et al.,  
664 2023). PGC retains nonpolar analytes through dispersive forces, while polar compounds are  
665 retained through charge induction on its polarizable surface. This stationary phase has been  
666 successfully used with TFA as an ion-pairing and acidifier agent. Furthermore, ion-pair HPLC  
667 with nonafluoropentanoic acid in pure water with 0.1% formic acid has been applied to  
668 separate 33 synthetic SCPs, revealing a positive correlation between retention time and  
669 hydrophobicity, particularly in relation to methylene elongation of amino acids. A notable



670 limitation of using ion-pairing agents, especially TFA, is their potential for permanent  
671 adsorption onto the PGC surface, leading to modifications of the stationary phase and  
672 necessitating regular regeneration.

673 The post-column addition of supercharging agent has been shown to enhance ion-source  
674 ionization of peptides (Abaye et al., 2021). Specifically, the use of 3-nitrobenzyl alcohol has  
675 demonstrated up to an 18-fold increase in ionization efficiency for standard SCPs, leading to  
676 the generation of higher charge states. However, no other supercharging agents have been  
677 tested for SCP ionization, leaving room for further exploration in this area.

678 The detection and fragmentation of SCPs require specific mass spectrometer settings. Due  
679 to their low molecular mass and distinct chemical structures, ion-source parameters must be  
680 optimized for the ionization of 'small' molecules. SCPs are predominantly detected as  
681 monocharged ions, which necessitates limiting the MS and MS/MS range to 1000 m/z to  
682 ensure optimal detection of both precursor and fragment ions. For MS/MS data acquisition,  
683 DDA mode has proven to be more effective than DIA mode when combined with an m/z  
684 inclusion list covering 4,980 unique SCP m/z ratios. The choice of fragmentation method is  
685 also critical, as SCPs have reduced chemical complexity compared to larger peptides. High-  
686 energy collision dissociation (HCD) and ultraviolet photodissociation (UV-PD) have emerged  
687 as the preferred fragmentation techniques, as they maximize the generation of peptide  
688 fragments. Immonium ions play a crucial role in SCP identification, serving as key diagnostic  
689 ions. Therefore, any MS fragmentation technique that efficiently produces immonium ions is  
690 highly recommended. The accuracy and speed of the MS analyzer directly influence the  
691 reliability of SCP identification, making advancements in high-resolution mass spectrometry  
692 particularly relevant. Despite these developments, recent innovations in mass spectrometry  
693 have yet to be fully explored in the context of SCP analysis. Notably, no study has  
694 investigated the potential of ion mobility mass spectrometry (IM-MS) to enhance SCP  
695 identification, highlighting an area for future research.

696 The identification of large- and medium-chain peptides is currently achieved through  
697 bioinformatic processing of MS/MS fragmentation data using specialized proteomics software  
698 in combination with protein databases. The peptide identification score generated by these  
699 tools ensures that the identification is statistically reliable, given the extensive MS/MS data  
700 available for longer peptides. In contrast, identification of SCPs presents a significant  
701 challenge due to the limited number of MS/MS fragmentation data points available for these  
702 shorter peptides. This limitation makes it impossible to achieve a robust identification score  
703 that guarantees the result is not due to chance. As a result, the identification SCPs must rely  
704 on *de novo* sequencing rather than database-dependent approaches. While manual *de novo*

705 sequencing is impractical for large-scale studies, it remains valuable for validating specific  
706 identifications.

707 A wide range of *de novo* sequencing algorithms has been developed, as recently reviewed  
708 by (Ng et al., 2023). These include graph-based, dynamic programming, and machine  
709 learning or neural network-based algorithms. However, none of these algorithms were  
710 specifically designed for SCP identification. Given the unique structural characteristics and  
711 analytical challenges associated with SCPs, there is an urgent need to develop dedicated *de*  
712 *novo* sequencing algorithms optimized for their identification.

713 Publishing the identity of peptides and proteins in high-quality scientific journals now requires  
714 the deposition of MS/MS raw data or identification data in online repositories. This practice  
715 allows for thorough review, consultation, and integration into big data studies. The  
716 identification of SCPs should adhere to the same standard for transparency and  
717 reproducibility. By depositing this data in accessible repositories, researchers enable others  
718 to validate, reuse, and build upon their work, promoting scientific collaboration and advancing  
719 the field.

720

## 721 **Acknowledgements**

722 This review was written within the framework of: **i)** the industrial chair ProteinoPeps, funded  
723 by INRAe and the French state; **ii)** the Charles Viollette industrial chair, supported by the  
724 MEL, the I-SITE, and the IAR; **iii)** the University of Artois & the *Université du Littoral Côte*  
725 *d'Opale*; and **iv)** the BiHauts Eco de France CPER/FEDER 2021-2027 program, financed by  
726 the European Union, the French State, and the Hauts-de-France Region.

727

## 728 **Author contributions: CRediT**

729 Alexandre Weber: conceptualization, visualization, writing-original draft. Barbara Deracinois:  
730 conceptualization, visualization, writing-original draft. Benoit Cudennec: funding acquisition,  
731 writing review & editing. Maxime de Sousa Lopes Moreira: visualization, writing-original draft.  
732 Rozenn Ravallec: funding acquisition, writing review & editing. Thierry Grard:  
733 conceptualization, project administration, supervision, writing review & editing. Renato  
734 Froidevaux and François Coutte: supervision. Christophe Flahaut: conceptualization, project  
735 administration, supervision, writing review & editing.

736

## 737 **References**

- 738 Abaye, D. A., Agbo Ab, I. A., & Nielsen, B. V. (2021). *Current perspectives on supercharging*  
739 *reagents in electrospray ionization mass spectrometry.*  
740 <https://doi.org/10.1039/d1ra00745a>
- 741 Aspevik, T., Oterhals, Å., Rønning, S. B., Altintzoglou, T., Wubshet, S. G., Gildberg, A.,  
742 Afseth, N. K., Whitaker, R. D., & Lindberg, D. (2017). Valorization of Proteins from Co-  
743 and By-Products from the Fish and Meat Industry. *Topics in Current Chemistry (Cham)*,  
744 375(3). <https://doi.org/10.1007/S41061-017-0143-6>
- 745 Boboua, S. Y. B., Wen, Q., Zhang, L., Chen, Y., Yu, J., Chen, P., Sun, Y., & Zheng, T.  
746 (2024). Valorization of animal waste proteins for agricultural, food production, and  
747 medicinal applications. *Frontiers in Sustainable Food Systems*, 8, 1366333.  
748 <https://doi.org/10.3389/FSUFS.2024.1366333/BIBTEX>
- 749 Bouglé, D., & Bouhallab, S. (2017). Dietary bioactive peptides: Human studies. *Critical*  
750 *Reviews in Food Science and Nutrition*, 57(2), 335–343.  
751 <https://doi.org/10.1080/10408398.2013.873766>
- 752 Cavaliere, C., Montone, A. M. I., Aita, S. E., Capparelli, R., Cerrato, A., Cuomo, P., Laganà,  
753 A., Montone, C. M., Piovesana, S., & Capriotti, A. L. (2021). Production and  
754 characterization of medium-sized and short antioxidant peptides from soy  
755 flour-simulated gastrointestinal hydrolysate. *Antioxidants*, 10(5), 734.  
756 <https://doi.org/10.3390/ANTIOX10050734/S1>
- 757 Cerrato, A., Aita, S. E., Capriotti, A. L., Cavaliere, C., Montone, A. M. I., Montone, C. M., &  
758 Laganà, A. (2022). Investigating the Short Peptidome Profile of Italian Dry-Cured Ham  
759 at Different Processing Times by High-Resolution Mass Spectrometry and  
760 Chemometrics. *International Journal of Molecular Sciences*, 23(6).  
761 <https://doi.org/10.3390/IJMS23063193/S1>
- 762 Cerrato, A., Aita, S. E., Capriotti, A. L., Cavaliere, C., Montone, C. M., Laganà, A., &  
763 Piovesana, S. (2020). A new opening for the tricky untargeted investigation of natural  
764 and modified short peptides. *Talanta*, 219(June), 121262.  
765 <https://doi.org/10.1016/j.talanta.2020.121262>
- 766 Cerrato, A., Aita, S. E., Cavaliere, C., Laganà, A., Montone, C. M., Piovesana, S., Zenezini  
767 Chiozzi, R., & Capriotti, A. L. (2021). Comprehensive identification of native medium-  
768 sized and short bioactive peptides in sea bass muscle. *Food Chemistry*, 343, 128443.  
769 <https://doi.org/10.1016/J.FOODCHEM.2020.128443>
- 770 Cerrato, A., Capriotti, A. L., Capuano, F., Cavaliere, C., Montone, A. M. I., Montone, C. M.,

771 Piovesana, S., Chiozzi, R. Z., & Laganà, A. (2020). Identification and Antimicrobial  
772 Activity of Medium-Sized and Short Peptides from Yellowfin Tuna (*Thunnus albacares*)  
773 Simulated Gastrointestinal Digestion. *Foods* 2020, Vol. 9, Page 1185, 9(9), 1185.  
774 <https://doi.org/10.3390/FOODS9091185>

775 Cittaro, D., Borsotti, D., Maiolica, A., Argenzio, E., & Rappsilber, J. (2005). Peptide  
776 Identification Using Vectors of Small Fragment Ions. *Journal of Proteome Research*, 4,  
777 1006–1011.

778 Du, A., & Jia, W. (2023). New insights into the bioaccessibility and metabolic fates of short-  
779 chain bioactive peptides in goat milk using the INFOGEST static digestion model and an  
780 improved data acquisition strategy. *Food Research International*, 169(May).  
781 <https://doi.org/10.1016/j.foodres.2023.112948>

782 Duca, F. A., Zaved Waise, T. M., Pepler, W. T., & Lam, T. K. T. (2021). The metabolic  
783 impact of small intestinal nutrient sensing. *Nature Communications*, 12(903).  
784 <https://doi.org/10.1038/s41467-021-21235-y>

785 Dugardin, C., Fleury, L., Touche, V., Ahdach, F., Lesage, J., Tenenbaum, M., Everaert, N.,  
786 Briand, O., Lestavel, S., Ravallec, R., & Cudennec, B. (2022). An Exploratory Study of  
787 the Role of Dietary Proteins in the Regulation of Intestinal Glucose Absorption. *Frontiers*  
788 *in Nutrition*, 8(January). <https://doi.org/10.3389/fnut.2021.769773>

789 Fleury, L., Deracinois, B., Dugardin, C., Nongonierma, A. B., FitzGerald, R. J., Flahaut, C.,  
790 Cudennec, B., & Ravallec, R. (2022). In Vivo and In Vitro Comparison of the DPP-IV  
791 Inhibitory Potential of Food Proteins from Different Origins after Gastrointestinal  
792 Digestion. *International Journal of Molecular Sciences*, 23(15).  
793 <https://doi.org/10.3390/ijms23158365>

794 Furey, A., Moriarty, M., Bane, V., Kinsella, B., & Lehane, M. (2013). Ion suppression; a  
795 critical review on causes, evaluation, prevention and applications. *Talanta*, 115, 104–  
796 122. <http://www.ncbi.nlm.nih.gov/pubmed/24054567>

797 Halligan, B. D., Dratz, E. A., Feng, X., Twigger, S. N., Tonellato, P. J., & Greene, A. S.  
798 (2004). Peptide Identification Using Peptide Amino Acid Attribute Vectors research  
799 articles. *Journal of Proteome Research*, 3(4), 813–820. [https://doi.org/DOI:](https://doi.org/DOI:10.1021/pr0499444)  
800 [10.1021/pr0499444](https://doi.org/DOI:10.1021/pr0499444)

801 Harscoat-Schiavo, C., Nioi, C., Ronat-Heit, E., Paris, C., Vanderesse, R., Fournier, F., &  
802 Marc, I. (2012). Hydrophilic properties as a new contribution for computer-aided  
803 identification of short peptides in complex mixtures. *Analytical and Bioanalytical*

804 *Chemistry*, 403(7), 1939–1949. <https://doi.org/10.1007/S00216-012-5987-6>

805 Heidenreich, E., Pfeffer, T., Kracke, T., Mechtel, N., Nawroth, P., Hoffmann, G. F., Schmitt,  
806 C. P., Hell, R., Poschet, G., & Peters, V. (2021). A novel uplc-ms/ms method identifies  
807 organ-specific dipeptide profiles. *International Journal of Molecular Sciences*, 22(18).  
808 <https://doi.org/10.3390/ijms22189979>

809 Heres, A., Gallego, M., Mora, L., & Toldrá, F. (2022). Identification and Quantitation of  
810 Bioactive and Taste-Related Dipeptides in Low-Salt Dry-Cured Ham. *International*  
811 *Journal of Molecular Sciences*, 23(5). <https://doi.org/10.3390/ijms23052507>

812 Hollebrands, B., Hageman, J. A., Sande, J. W. Van De, & Albada, B. (2023). Improved LC –  
813 MS identification of short homologous peptides using sequence - specific retention time  
814 predictors. *Analytical and Bioanalytical Chemistry*, 415(14), 2715–2726.  
815 <https://doi.org/10.1007/s00216-023-04670-2>

816 Huang, Y. P., Dias, F. F. G., Leite Nobrega de Moura Bell, J. M., & Barile, D. (2022). A  
817 complete workflow for discovering small bioactive peptides in foods by LC-MS/MS: A  
818 case study on almonds. *Food Chemistry*, 369.  
819 <https://doi.org/10.1016/J.FOODCHEM.2021.130834>

820 Karami, Z., & Akbari-adergani, B. (2019). Bioactive food derived peptides: a review on  
821 correlation between structure of bioactive peptides and their functional properties.  
822 *Journal of Food Science and Technology* 2019 56:2, 56(2), 535–547.  
823 <https://doi.org/10.1007/S13197-018-3549-4>

824 Lahrichi, S. L., Affolter, M., Zolezzi, I. S., & Panchaud, A. (2013). Food Peptidomics: Large  
825 scale analysis of small bioactive peptides - A pilot study. *Journal of Proteomics*, 88, 83–  
826 91. <https://doi.org/10.1016/j.jprot.2013.02.018>

827 Le Maux, S., Nongonierma, A. B., & Fitzgerald, R. J. (2015). Improved short peptide  
828 identification using HILIC–MS/MS: Retention time prediction model based on the impact  
829 of amino acid position in the peptide sequence. *Food Chemistry*, 173, 847–854.  
830 <https://doi.org/10.1016/J.FOODCHEM.2014.10.104>

831 Le Maux, S., Nongonierma, A. B., Murray, B., Kelly, P. M., & FitzGerald, R. J. (2015).  
832 Identification of short peptide sequences in the nanofiltration permeate of a bioactive  
833 whey protein hydrolysate. *Food Research International*, 77, 534–539.  
834 <https://doi.org/10.1016/J.FOODRES.2015.09.012>

835 Mann, B., Athira, S., Sharma, R., Kumar, R., & Sarkar, P. (2019). Bioactive Peptides from  
836 Whey Proteins. *Whey Proteins*, 519–547. <https://doi.org/10.1016/B978-0-12-812124->

837 5.00015-1

838 Montone, C. M., Capriotti, A. L., Cerrato, A., Antonelli, M., La Barbera, G., Piovesana, S.,  
839 Laganà, A., & Cavaliere, C. (2019). Identification of bioactive short peptides in cow milk  
840 by high-performance liquid chromatography on C18 and porous graphitic carbon  
841 coupled to high-resolution mass spectrometry. *Analytical and Bioanalytical Chemistry*,  
842 *411*(15), 3395–3404. <https://doi.org/10.1007/S00216-019-01815-0>

843 Ng, C. C. A., Zhou, Y., & Yao, Z. P. (2023). Algorithms for de-novo sequencing of peptides  
844 by tandem mass spectrometry: A review. *Analytica Chimica Acta*, *1268*, 341330.  
845 <https://doi.org/10.1016/J.ACA.2023.341330>

846 Nongonierma, A. B., Gaudel, C., Murray, B. A., Flynn, S., Kelly, P. M., Newsholme, P., &  
847 FitzGerald, R. J. (2013). Insulinotropic properties of whey protein hydrolysates and  
848 impact of peptide fractionation on insulinotropic response. *International Dairy Journal*,  
849 *32*(2), 163–168. <https://doi.org/10.1016/J.IDAIRYJ.2013.05.014>

850 O'keeffe, M. B., & Fitzgerald, R. J. (2015). *Identification of short peptide sequences in*  
851 *complex milk protein hydrolysates*. <https://doi.org/10.1016/j.foodchem.2015.03.077>

852 Panchaud, A., Affolter, M., & Kussmann, M. (2012). Mass spectrometry for nutritional  
853 peptidomics: How to analyze food bioactives and their health effects. *Journal of*  
854 *Proteomics*, *75*(12), 3546–3559. <https://doi.org/10.1016/J.JPROT.2011.12.022>

855 Piovesana, S., Capriotti, A. L., Cerrato, A., Crescenzi, C., Barbera, G. La, Laganà, A.,  
856 Montone, C. M., & Cavaliere, C. (2019). Graphitized Carbon Black Enrichment and  
857 UHPLC-MS/MS Allow to Meet the Challenge of Small Chain Peptidomics in Urine.  
858 *Analytical Chemistry*, *91*(17), 11474–11481.  
859 <https://doi.org/10.1021/ACS.ANALCHEM.9B03034>

860 Piovesana, S., Cerrato, A., Antonelli, M., Benedetti, B., Capriotti, A. L., Cavaliere, C.,  
861 Montone, C. M., & Laganà, A. (2020). A clean-up strategy for identification of circulating  
862 endogenous short peptides in human plasma by zwitterionic hydrophilic liquid  
863 chromatography and untargeted peptidomics identification. *Journal of Chromatography*  
864 *A*, *1613*. <https://doi.org/10.1016/J.CHROMA.2019.460699>

865 Piovesana, S., Laganà, A., & Capriotti, A. L. (2023). Current advances in carbonaceous  
866 materials for analytical applications in liquid phase. *TrAC Trends in Analytical*  
867 *Chemistry*, *168*, 117297. <https://doi.org/10.1016/J.TRAC.2023.117297>

868 Piovesana, S., Montone, C. M., Cavaliere, C., Crescenzi, C., La Barbera, G., Laganà, A., &  
869 Capriotti, A. L. (2019). Sensitive untargeted identification of short hydrophilic peptides

870 by high performance liquid chromatography on porous graphitic carbon coupled to high  
871 resolution mass spectrometry. *Journal of Chromatography A*, 1590, 73–79.  
872 <https://doi.org/10.1016/j.chroma.2018.12.066>

873 Pripp, A. H. (2007). Docking and virtual screening of ACE inhibitory dipeptides. *European*  
874 *Food Research and Technology*, 225(3–4), 589–592. [https://doi.org/10.1007/S00217-](https://doi.org/10.1007/S00217-006-0450-6/FIGURES/2)  
875 [006-0450-6/FIGURES/2](https://doi.org/10.1007/S00217-006-0450-6/FIGURES/2)

876 Takano, Y., Kashiya, Y., Ogawa, N. O., Chikaraishi, Y., & Ohkouchi, N. (2010). Isolation  
877 and desalting with cation-exchange chromatography for compound-specific nitrogen  
878 isotope analysis of amino acids: Application to biogeochemical samples. *Rapid*  
879 *Communications in Mass Spectrometry*, 24(16), 2317–2323.  
880 <https://doi.org/10.1002/RCM.4651>

881 Takano, Y., Oba, Y., Furota, S., Naraoka, H., Ogawa, N. O., Blattmann, T. M., & Ohkouchi,  
882 N. (2021). Analytical development of seamless procedures on cation-exchange  
883 chromatography and ion-pair chromatography with high-precision mass spectrometry for  
884 short-chain peptides. *International Journal of Mass Spectrometry*, 463, 116529.  
885 <https://doi.org/10.1016/j.ijms.2021.116529>

886 Yesmine, B. H., Antoine, B., da Silva Ortência Leocádia, N. G., Rogério, B. W., Ingrid, A.,  
887 Nicolas, B., Thierry, M., Piot, J. M., Frédéric, S., & Stéphanie, B. J. (2017). Identification  
888 of ace inhibitory cryptides in Tilapia protein hydrolysate by UPLC–MS/MS coupled to  
889 database analysis. *Journal of Chromatography B: Analytical Technologies in the*  
890 *Biomedical and Life Sciences*, 1052, 43–50.  
891 <https://doi.org/10.1016/j.jchromb.2017.02.015>

892