LWT - Food Science and Technology 73 (2016) 296–302

Contents lists available at ScienceDirect

LWT - Food Science and Technology

journal homepage: www.elsevier.com/locate/lwt

Optimal computational comparison of mass spectrometric peptide profiles of alternative hydrolysates from the same starting material

Thérèse A. Holton [a, b, c], Eugène T. Dillon [a, b, d], Aisling Robinson [a, b, d], Kieran Wynne [b, e], Gerard Cagney [a, b, d], Denis C. Shields [a, b, c, *]

[a] Food for Health Ireland, University College Dublin, Dublin 4, Ireland
[b] UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland
[c] School of Medicine, University College Dublin, Dublin 4, Ireland
[d] School of Biomolecular & Biomedical Science, University College Dublin, Dublin 4, Ireland
[e] Mass Spectrometry Resource, University College Dublin, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form
9 February 2016
Accepted 8 June 2016
Available online 9 June 2016

Keywords:
Bioactive peptides
Characterisation
Cleavage patterns
Hydrolysates
Peptidomics

ABSTRACT

Food protein hydrolysates are complex mixtures that are increasingly being analysed by tandem mass spectrometry. Given a single starting material, many alternative peptide profiles are achievable under varying hydrolysis conditions. To date, characterisation of the relative similarities and differences between such peptide profiles remains largely unstudied. Here, we investigate optimal computational methods for grouping peptide profiles of hydrolysates derived from the same starting material. Using an experimental bovine milk dataset, we evaluated how these methods grouped either technical replicates, or distinct samples with known cleavage profiles. Analyses performed using only the presence and abundance of peptides were found to be suboptimal for achieving effective sample grouping. Using the amino acid distribution at both termini of peptides was more efficient at grouping replicate samples; however, this approach lacked suitable discrimination between distinct samples. By extending the termini approach to incorporate the abundance associated with terminal amino acids, optimal grouping was achieved. We therefore suggest that grouping peptide profiles of hydrolysates from the same starting material should rely on a combination of N and C terminal amino acid frequency and abundance. Importantly, this approach requires no *a priori* knowledge of enzyme specificities, making it generally applicable to diverse sets of food matrices.

1. Introduction

Protein hydrolysis is a process that is routinely carried out in food production settings and has a long established role in improving the functional and nutritional qualities of foods (Neklyudov, Ivankin, & Berdutina, 2000; Panyam & Kilara, 1996). With growing commercial interest in functional foods, nutraceuticals and supplementation, the use of protein hydrolysis is increasingly extending to the production and release of bioactive peptides (Phelan, Aherne, FitzGerald, & O'Brien, 2009). Traditionally, industry scale protein hydrolysis was achieved by microbial fermentation or by chemical hydrolysis. However, owing to their greater specificity and reduced impact on environmental conditions (Tavano, 2013), there has been a continued progression towards the use of enzymes in industry settings since the 1960s (Fernandes, 2010).

Characterisation and profiling of protein hydrolysates is imperative both from a production and functional perspective. In particular, it is often advantageous to compare samples in order to monitor batch-to-batch variation or to identify similar or indeed, disparate features between hydrolysates. Owing to the typically complex composition of hydrolysates, characterisation can be difficult and is generally achieved through the assessment of various physiochemical properties. Characterisation approaches routinely employed include assays for determining the degree of hydrolysis (DH) and chromatography techniques such as RP-HPLC and size exclusion chromatography (SEC), that separate based on solubility and size respectively (Silvestre, 1997). Additionally,

Abbreviations: DH, Degree of Hydrolysis; SEC, Size Exclusion Chromatography; LFQ, Label Free Quantitative; PCA, Principal Components Analysis.

* Corresponding author. UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland.
E-mail address: denis.shields@ucd.ie (D.C. Shields).

http://dx.doi.org/10.1016/j.lwt.2016.06.025
0023-6438/

various mass spectrometry (MS) techniques are used for hydrolysate characterisation including FAB, ESI and MALDI (Léonil, Gagnaire, Mollé, Pezennec, & Bouhallab, 2000).

The advent of the peptidomics era has introduced the capacity to elucidate the full complement of native peptides (be it endogenous peptides, or those created through hydrolysis or fermentation) in samples of interest. Indeed, continued advances in tandem MS technologies provide for progressively more accurate identification and quantification of such peptides (Michalski et al., 2011). Accordingly, this technology has attracted the attention of food and nutritional researchers as a promising approach for the characterisation of food hydrolysates (Carrasco-Castilla, Hernández-Álvarez, Jiménez-Martínez, Gutiérrez-López, & Dávila-Ortiz, 2012; Dallas et al., 2015; Ibáñez, Simó, García-Cañas, Cifuentes, & Castro-Puyana, 2013; Lahrichi, Affolter, Zolezzi, & Panchaud, 2013; Panchaud, Affolter, & Kussmann, 2012). While the high concentration of low molecular weight peptides in food hydrolysates and fermentates undoubtedly presents an analytical challenge to peptidomic analyses (see Panchaud et al., 2012), research in this area is yielding significant progress with the anticipation that continued progression will allow for the identification of all small peptides (Lahrichi et al., 2013).

Although the utility of peptidomics in food and nutrition research is apparent, the volume of data produced by such techniques presents a challenge in terms of appropriate data analysis (Cappadona, Baker, Cutillas, Heck, & van Breukelen, 2012). For many years, bioinformatics has facilitated various aspects of proteomic analyses (Palagi, Hernandez, Walther, & Appel, 2006) across many disciplines including food science and nutrition (Holton, Vijayakumar, & Khaldi, 2013; Ibáñez et al., 2013), and is now playing a similar role in peptidomic analyses (Menschaert et al., 2010). To date, development of computational methods for sample comparison has been driven by the identification of biomarkers (Ling et al., 2010; Nanni et al., 2009; Nordén, Broberg, Lindberg, & Plymoth, 2005; Ueda et al., 2011) and post-translational modifications (Lahrichi et al., 2013; Lundby, Lage, et al., 2012; Lundby, Secher, et al., 2012). Such sample comparison methods often use the peaks in the mass spectra (Key, 2012; Ling et al., 2010; Nanni et al., 2009; Schmidt, McIlwain, Page, Christie, & Li, 2008) to differentiate between samples, however, recently there has been a progression towards the use of peptide sequences (Lambers et al., 2015; Wang et al., 2012) and quantitative data (Liu et al., 2012; Nagaraj & Mann, 2011).

We set out to optimise the computational approaches for differentiating quantitative peptidomic profiles from a single food starting material. As a test dataset, we chose exemplar hydrolysates of bovine milk digested using various enzymes with discrete and well-characterised hydrolysis patterns. Our selected enzymes included ArgC and LysC, which cleave at the C termini of arginine and lysine respectively, and trypsin, which cleaves at both aforementioned sites. To reflect the complexity of enzyme mixtures, we additionally included an ArgC and LysC (ArgC-LysC) combination hydrolysate, which is expected to converge on the cleavage pattern of trypsin. We found that abundance weighting and confining analyses to peptide termini provided the optimal computational approach for grouping hydrolysates derived from the same starting material. This approach can have widespread application in food science and nutrition, wherever samples of the same starting material need to be grouped or classified according to their peptide profiles, either in research, development or in the analysis of batch-to-batch variation during production.

2. Materials and methods

2.1. Sample preparation and mass spectrometry analysis

Skim milk powder (Sigma Aldrich 70166) was resuspended at 2% w/v in deionized water and diluted to a concentration of 1 ug/ul. Disulfide bonds were reduced using dithiothreitol followed by alkylation with iodoacetamide. Digestion with sequencing grade bovine trypsin (Abnova P5320), porcine trypsin (Sigma Aldrich T6567), ArgC (Sigma Aldrich P6056) or LysC (Sigma Aldrich P3428) was carried out overnight at 37 °C. The enzyme to protein concentration for all hydrolysates was 1:100. To discount the influence of endogenous milk peptides on the computational comparisons, a control sample of non-enzyme treated skim milk was prepared and analysed analogously to the enzyme-digested samples. Digestion was stopped with the addition of trifluoroacetic acid (TFA), peptides were desalted with C18 STAGE tips (Rappsilber, Mann, & Ishihama, 2007) and resuspended in 0.1% TFA.

All samples were run in triplicate (yielding three technical replicates per sample) on a Thermo Scientific Q Exactive mass spectrometer connected to a Dionex Ultimate 3000 (RSLCnano) chromatography system. Each sample was loaded onto a fused silica emitter (75 μm ID, pulled using a laser puller (Sutter Instruments P2000)), packed with Reprocil Pur C18 (1.9 μm) reverse phase media and was separated by an increasing acetonitrile gradient over 47 or 59.5 min at a flow rate of 250 nL/min. The mass spectrometer was operated in positive ion mode with a capillary temperature of 320 °C, and with a potential of 2300 V applied to the frit. All data was acquired with the mass spectrometer operating in automatic data dependent switching mode. A high resolution (70,000) MS scan (300–1600 m/z) was performed using the Q Exactive to select the 12 most intense ions prior to MS/MS analysis using HCD.

Raw data from the Q-Exactive was processed using MaxQuant version 1.5.1.0 (Cox & Mann, 2008), incorporating the Andromeda search engine (Cox et al., 2011). To identify peptides and proteins, MS/MS spectra were searched against a custom bovine milk database using the default settings of MaxQuant. All searches were performed in unspecific digestion mode allowing two missed cleavages. The database searches were performed with carbamidomethylation (C) as a fixed modification and acetylation (protein N terminus) and oxidation (M) as variable modifications. For the generation of label free quantitative (LFQ) ion intensities for protein profiles, signals of corresponding peptides in different LC-MS/MS runs were matched by MaxQuant (Cox et al., 2014).

2.2. Peptide sequence and abundance data assembly

For each of our six samples (five enzyme preparations and an unhydrolysed bovine milk control), we considered the peptide sequence and their corresponding abundance profiles returned by MaxQuant (Cox & Mann, 2008). Both replicate data and a single consensus profile per sample (achieved by dividing the summed triplicate intensities for each peptide by three) were subject to analysis (see Supplementary Materials for a list of peptides identified in each consensus sample). In-house Perl scripts were used to extract the N and C terminal, and theoretical N-1 and C+1, amino acid frequencies in each sample. Here, N-1 and C+1 amino acids are those that are present in the parent protein sequence to the left and right respectively of the N and C term end of each peptide. Thus, this data represents cleavage not seen in the set of characterised MS/MS peptides but that was inferred from adjacent residues in the parent protein sequence. Where required, the abundance of the source peptide was additionally incorporated into these amino acid frequencies. This was achieved by multiplying all N and C terminal frequencies by their source peptide abundances. The resulting values were then normalised by dividing by the total abundance of all peptides in the respective sample, however, analyses conducted precluding this step returned analogous results (data not shown). Amino acids at the termini of proteins were excluded from all