Journal of Food Composition and Analysis 73 (2018) 83–90

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

Original research article

Peptide biomarkers for identifying the species origin of gelatin using coupled UPLC-MS/MS

Sean Ward, Nicholas T. Powles, Michael I. Page*

IPOS, The Page Laboratories, Department of Chemical Sciences, The University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ARTICLE INFO

Keywords:
Food analysis
Food composition
Gelatin
Species origin
LC–MS/MS
Proteomics
MPP

ABSTRACT

Liquid chromatography linked with mass spectrometry (LC–MS/MS) was used to analyse gelatin from four different species after a trypsin digest. Using chemometric software to analyse the data it was possible to find peptide fragments that were specific to each species of gelatin: porcine, bovine, chicken or fish. Identification of these peptides was challenging due to the destructive nature of gelatin manufacture. The untargeted workflow method developed allowed identification of 21 unknown gelatin samples with 100% accuracy. Fish gelatin is made from a large range of different species that do not share a common differentiating protein but it was shown that the protein from a parasitic bacteria could be used to identify fish gelatin.

1. Introduction

Gelatin is a mixture of polypeptides produced by the partial hydrolysis of the connective tissue protein collagen recovered from the bones and hides of animals, mainly bovine and porcine (Venien and Levieux, 2005). It is an important product, which has many applications in the food industry, particularly as a gelling agent, stabiliser and thickener, as well as in the medical and cosmetic industries (Zhang et al., 2009). Collagen, which is present in all multicellular organisms, is not one protein but a family of structurally related ones. The different collagen proteins have very diverse functions and are characterised by different polypeptide compositions but all refer to a group of extracellular matrix proteins composed of a triple α−helical structure (Brodsky and Ramshaw, 1997). The three chains may be all identical or may be of two different chains – two identical α_1 chains and one α_2 chain with a slightly different amino acid sequence (Maynes, 1987). In higher animals there are at least 19 different types of collagen known to exist. The trimeric form of type I collagen may have a molecular mass of ~400 kDa. There are two unique features associated with the structure of collagen – every third amino acid is usually glycine which gives a repeat of $(Gly\text{-}X\text{-}Y)_n$ and about 20% of the amino acids are either proline or hydroxyproline (Brodsky and Ramshaw, 1997).

Gelatin is produced from any collagen containing tissue, although the hides, skin and bones of porcine or bovine animals are preferred but some gelatin is produced from fish and fowl. The manufacturing process starts with the cleaning of the raw tissue, acid or alkaline hydrolysis followed by the extraction, purification and neutralisation processes (Phillips and Williams, 2009). Acid pre-treatment produces type A gelatins and type B gelatin by an alkaline pre-treatment. It is the combination of pre-treatment and extraction processes, which gives the final product its distribution of polypeptide chains with differing molecular weights. The outbreak of bovine spongiform encephalopathy (BSE) in 1986 led to restrictions being implemented by the regulatory authorities on the use of bovine gelatin for human consumption. There are also restrictions by some religions and cultures banning the consumption of porcine products and so it has become necessary to develop a simple but rigorous analytical method to determine the species of origin of gelatin (Venien and Levieux, 2005).

The aim of this work was to develop a robust mass spectrometric method for the biomarker identification of the species origin of gelatin samples. An unsuccessful solvent extraction of low molecular weight species from gelatin was initially attempted in order to identify possible biomarkers. Subsequently a proteomics (Anderson and Anderson, 1998) method was used. There are two approaches to the proteomic workflow, either a comprehensive “shotgun” approach or a targeted method (Orton and Doucette, 2013; Darie, 2013; Yang et al., 2018). The shotgun approach looks at all of the proteins in a complex mixture, whereas the targeted approach only looks for specific proteins. In a targeted approach the sample preparation and MS conditions are optimised for specific proteins (Mitchell, 2010). Preparation of the sample for proteomics analysis is an important step with a proteolytic enzyme used to cleave the protein into peptides, which are in a suitable mass range for MS analysis (Steen and Mann, 2004). The most commonly used digestive enzyme in MS proteomics is trypsin, the high specificity

* Corresponding author.
E-mail address: m.i.page@hud.ac.uk (M.I. Page).

https://doi.org/10.1016/j.jfca.2018.08.002
Received 15 May 2018; Received in revised form 24 July 2018; Accepted 15 August 2018
Available online 18 August 2018
0889-1575/

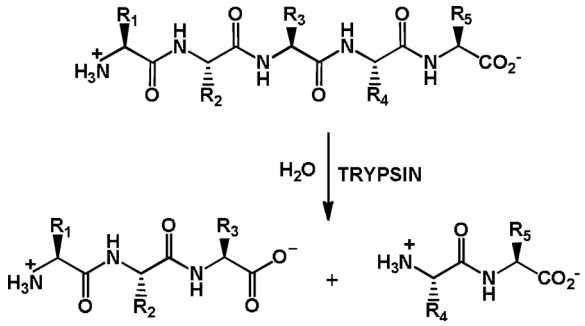

Fig. 1. Tryptic cleavage of peptide bond where R_1 and R_2 are any amino acids, R_3 is lysine or arginine and R_4 is not proline.

of which aids database matching (Olsen et al., 2004). Trypsin preferentially cleaves the peptide bond on the carboxylic acid side of arginine or lysine, except when followed by proline (Rodriguez et al., 2008) (Fig. 1). This generates a basic fragment from the original *N*-terminus end of the peptide together with at least one protonated lysine or arginine at the new *C*-terminus; so tryptic digest peptides are expected to have a charge state of at least 2+ when analysed by MS using electrospray ionisation, assuming that all of the carboxylates are protonated to neutral carboxylic acids. The ionisation efficiency of electrospray ionisation (ESI) is highly dependent on both the pH of the mobile phase and the pK_a value of the analyte (Naegele, 2011). Acidic solutions of 0.1% formic acid are often used in positive ion mode, giving the mobile phase a pH of 2.7 which is about one unit below the pK_a of the *C*-terminal carboxylic acid. Furthermore, there is an excess of protons in the ionisation source for protonating the available basic sites (Liu et al., 2011).

In practice, there are often contaminating proteolytic enzymes in trypsin digests which will act on proteins in an unpredictable manner. Therefore, the concept of "semi-tryptic" digests has been developed to compensate for this additional cleavage activity. This cleavage type assumes that in addition to the major cleavage site ([KR]|{P}), other sites may be at any residue, which has led to many databases allowing non-tryptic or half-tryptic peptide matches (Alves et al., 2008; Olsen et al., 2004). The detection of 'non-tryptic' peptides in trypsin digests could also be due to fragmentation of the peptides in the ionisation source of the MS (Steen and Mann, 2004) – the protonated peptides may fragment along the peptide backbone (Medzihradszky and Chalkley, 2013) and even the side-chain (Roepstorff and Fohlman, 1984). The mass spectra produced by the chromatographic separation of trypsin-digested proteins lead to the production of very complex data sets. There can be thousands of peptides detected, which leads to the need for sophisticated software for data interpretation (Krenkova et al., 2009). The identification of different peptides from their fragmentation pattern may be possible by searching databases such as Mascot, Sequest or Phenyx (Cottrell, 2011). For uncharacterised peptides *de-novo* sequencing can be performed using software such as Peaks.

2. Materials and methods

2.1. Samples, reagents and software

Bovine and porcine gelatin, trypsin and buffer salts were purchased from Sigma Aldrich, UK. Three different certified samples of pork, beef, chicken and fish gelatin were obtained from different suppliers or different batches; 21 certified samples of unknown gelatin were supplied by Healan Ingredients Ltd., York, UK. LC/MS grade acetonitrile, Optima and LC–MS grade methanol were from Fisher Scientific, Loughborough, UK.

MassHunter Profinder software (Profinder) is a standalone software program which has been developed for batch feature extraction and alignment of TOF and Q-TOF data from GC, LC and CE instruments along with nominal mass GC/MS systems (Agilent G3835AA MassHunter Profinder Software, 2013). MassHunter Qualitative analysis (MassHunter Qual) software uses the "molecular feature extraction" tool to extract the feature combining the different charge states, isotopes and adducts, grouping them together and assigning a neutral mass. A compound chromatogram is then created for the compound which is a sum of all of the ions associated with it (Sana, 2017). Mass Profiler Professional (MPP) software is a powerful chemometrics platform, which has been designed to deal with highly complex MS data. MPP can be used in any MS-based differential analysis containing two or more sample groups or variables. It has both statistical analysis tools and visualisation tools. MPP is compatible with data from GC/MS, LC/MS, CE/MS and ICP/MS, allowing integrated identification or annotation of compounds along with pathway analysis for metabolomics and proteomic studies. MPP also enables automated sample class prediction for qualitative analysis of unknown samples in many different applications (Mass Profiler Professional Software, 2017). PEAKS is a proteomics software program for tandem mass spectrometry, designed for peptide sequencing, protein identification and quantification (Bioinformatics Solutions Inc., Waterloo, Ontario, Canada). The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data (https://www.uniprot.org).

2.2. Solvent extraction

One g of bovine and porcine gelatin was stirred at room temperature for thirty minutes in either 100 mL of methanol or acetonitrile. The samples were then syringe filtered and analysed by LC–MS (Table 1 Supplementary data). The data was extracted using Agilent Technologies Profinder software. Only species with an MS peak height greater than 5000 counts were included and alignment parameters of 10 ppm mass window and retention time drift of 0.5 min. All other settings were left at default. To determine the optimum conditions for extraction 1 g of beef or porcine gelatin was stirred and refluxed at 50, 60 and 70 °C in 100 mL of methanol. The experiments were sampled both initially when the temperature was reached and after 5 min. Each sample was studied in triplicate. The samples were then syringe filtered and analysed by LC–MS. The greatest responses were when the gelatin samples were extracted after 5 min at 70 °C. After extraction compound exchange format (CEF) files were generated. The CEF files were then uploaded into Mass Profiler Professional (MPP) software with the same alignment settings as used in Profinder. The results were analysed by principle component analysis (PCA) using Agilent Technologies MPP software (Supplementary data Fig. 1).

2.3. Trypsin digest

One hundred milligrams of each gelatin sample were added to 50 mL of ammonium bicarbonate solution (1% *w/v*, pH 8.0) and heated to 90 °C with stirring to give a clear solution, which was then filtered through a 0.22-μ m syringe filter. To 1.0 mL of this filtrate was added 50 μL of 2.5% trypsin solution and this mixture was incubated for 12 h at 37 °C. The samples were then analysed by LC–MS and the data extracted and analysed with Profinder and MPP, as previously described for the solvent extraction methodology.

To determine the optimum time for enzyme digestion, samples of pork, beef and chicken gelatin were prepared as for the trypsin digestion above, using an autosampler at 37 °C and injections carried out every hour for 12 h. The majority of changes occurred within the first hour but the data points for each species continued to diverge, showing a maximum benefit of a 12-hour digestion. To improve the separation of the peptides found in each sample and to reduce the number of co-eluting peptides, samples were prepared from trypsin-digested samples. A basic HPLC method was used to analyse the sample and the data examined using MassHunter Qualitative analysis software. Changes