Original research article

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

Sean Ward,[1] Nicholas T. Powles[1] and Michael I. Page[1]

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

Email: m.i.page@hud.ac.uk

Abstract

Liquid chromatography linked with mass spectrometry (LC-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.

Keywords Food analysis, food composition, gelatin, species origin, LC-MS, proteomics, MPP.

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 characterized by different polypeptide compositions but all refer to a group of extra-cellular matrix proteins composed of a triple α–helical structure (Brodsky et al, 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 ~400kDa. 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 et al, 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 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 which 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 for 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 digestions 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 leads 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 on 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 hplc-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 Software (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 visualization 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 gram 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 a MS peak height greater than 5,000 counts were included and alignment parameters of 10ppm 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 gram of beef or porcine gelatin was stirred and refluxed at 50, 60 and 70^{o}C in 100 mL of methanol. The experiments were sampled both initially when the temperature was reached and after 5 min.. Each sample was carried out 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^{o}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

100 mg of each gelatin sample was added to 50 mL of ammonium bicarbonate solution (1%$^{w}/_{v}$) pH 8.0 and heated to 90^{o}C with stirring to give a clear solution which was then filtered through a 0.22 μ syringe filter. To 1.0 mL of this filtrate was added 50 μL of 2.5% trypsin solution and incubated for 12 hrs at 37^{o}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 auto sampler at 37^{o}C and injections carried out every hour for 12 hrs. The majority of changes occurred within the first hour but the data points for each species continue to diverge showing a maximum benefit of a 12 hrs. 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 were then made to the LC gradient table giving isocratic regions at 10% and 15% acetonitrile - these two different gradients were then run with flow rates of 0.4 mL/min and 0.25 mL/min. Some samples used an injection volume of 10 µL rather than the previously used 5 µL. (Supplementary data Tables 2-4 Figs. 2-5).

2.4 Identification of peptide biomarkers for differentiating Pork, Beef, Chicken and Fish gelatin

Three different certified samples of pork, beef chicken and fish gelatin were obtained from different suppliers and different batches and digested with trypsin for 12 hrs and the digests analysed by LC-MS using the HPLC conditions developed (Supplementary data). and each sample was analysed in triplicate. The data was prepared for MPP analysis by extracting the compounds with Profinder, the first and last quality controlled replicate injections were opened in Masshunter Qual and showed good overlay (Supplementary data Fig. 6). The retention time cut-off was assessed by overlaying the QC runs with a system blank (**Fig. 2**) which showed few compounds eluting before 1min. The total ion counts for the blank and QC runs were similar after 21 min. so a retention time window of 1-21 min. was used for Profinder. At 21 min. the solvent ratio was 45% acetonitrile and 55% water, indicating that all the compounds of interest had eluted before the solvent gradient reaches 45% organic. The noise level was assessed in the solvent front region of the spectrum (before 1 min.) and was around 300 counts and so for Profinder extraction a value of 900 counts was used.

The parameters used in the project wizard for data extraction with Profinder software were a maximum charge state 3+, this being the highest charge state seen in the raw data for the digested peptides based on isotope spacing. A retention time window of 0.1 min. was used as determined from the retention time drift seen by comparing the first and last QC runs in Qualitative analysis. Next a minimum height threshold cut off of 5,000 counts was set and only allowing selection if they appeared in all three sample replicates to give a large number of entities to pass the screening process.

A larger ppm window for recursive data mining of 35 ppm was used using a symmetric window and a retention time window of 1.5 min. was also used. The matching scores were weighted at 100% mass accuracy, 60% isotope abundance and 50% isotope score. Retention time was set at 0% to ensure the entities were not seen just outside the retention time window used for initial filtering. Integration was left at the default of agile without smoothing and a peak height cut-off of 1,000 counts was used. Finally, the filter used was to limit the results to the largest 2,000 compounds to lower data processing time.

2.5 Synthesis of gelatin under controlled conditions

Finding reliable sources of uncontaminated gelatin which has fully traceable origins is not easy, so pure species specific gelatin samples were prepared. Beef, pork, fish and chicken gelatin skins were obtained from slaughter houses and washed in warm water to remove all of the blood and loose tissue. The hair was removed from the beef hide by means of mechanical scraping. The animal hides were then hydrolysed at room temperature with 4 weight equivalents of 5% HCl. The beef hide was hydrolysed for 5 days and the others for 24 hrs. Beef hide was hydrolysed for longer due it having a higher amount of collagen cross linking. After the hydrolysis the hides were washed in a large amount of water before 4 weight equivalents of water being added and the pH adjusted to pH 7 with dilute sodium hydroxide. The soaked hide was then heated to 90ºC for 3 hrs. A fat layer formed on the top and was discarded, the remaining liquid was then filtered and trypsin digested as before.

3. Results and Discussion

3.1 Solvent extraction

The aim of solvent extraction was to determine if either methanol or acetonitrile could be used to obtain biomarkers from various sources of gelatin. The methanol and acetonitrile extractions of pork gelatin yielded 157 and 120 compounds, respectively. For beef gelatin the corresponding yields were 126 and 129 compounds. However, no useful separation of the constituents to identify the species origin of the gelatin was evident.

3.2 Trypsin digest

Partial hydrolysis of the gelatin samples using trypsin were undertaken to determine if the resulting smaller peptides could be used to find biomarkers from different species. The PCA plot (**Fig. 3**) shows beef (red) and porcine (blue) samples are well separated by the digested process. Although there were no significant differentiators found with solvent extraction of the two species of gelatin, they can be separated by the different peptide sequences found in the digested samples. The difference between the two gelatin species falls along the principle component, this describes more of the variation in the data (27%) than the difference between the samples from the same species which is distributed along the second component (16%). The PCA plot shows that species identification using peptide identifiers may be possible.

Due to the large number of peptide fragments present in the sample and in the absence of known targets complete separation of all of the peptides proved difficult. The best separation was with an isocratic portion of the gradient starting at 10% acetonitrile and a flow rate of 0.25 mL/min with many of the peptides eluting in the large fractions from 6.5-10 min.

3.3 Identification of peptide biomarkers for differentiating Pork, Beef, Chicken and Fish gelatin

The CEF files generated in Profinder were input into MPP to look for unidentified compounds with an analysis type of significance testing and fold change. The data type selected was Masshunter Qual., a minimum abundance of 5,000 counts was used and a minimum number of ions of 2 and all charge states were permitted. A retention time window of 0.15 minutes was used along with a 15 ppm mass window. At this stage 304 compounds were seen in all 15 samples, 172 in 12 of the samples, 100 in 9 of the samples, 295 in 6 of the samples and 27 seen in 3 of the samples. Next the compounds were filtered by frequency, for this the compounds had to be present in 100% of samples in 1 sample group which filtered out only 7 compounds. The PCA plot in the guided workflow (**Fig. 4**) shows good species grouping (red = beef, blue = chicken, brown = fish and grey = pork) with really good reproducibility in the QC samples (green). This demonstrates that the analysis is reproducible and differences in the QC samples are smaller than the species variation.

After the guided workflow MPP was used as a filter to find the compounds that had only been detected in one particular species of each gelatin sample. Interpretations were created for pork, beef, fish and chicken gelatin and were based on species type and were numerical to allow lines between compound groups to be connected on the filtering plots. The average over the replicates was used with flags for present and marginal compounds. For each of the interpretations an entity list was created using filter by flags. For both of these only the present/marginal compounds were used. For each species an interpretation with the compounds being present in 3 out of 3 samples was made.

A Venn diagram was created using the filter by flag entities lists for each species (**Fig. 5**) showing those areas which do not overlap corresponding to compounds found only in one particular species : 86 only in pork, 23 in fish, 155 in chicken and 98 only in beef. These were combined to make a new entity list containing 362 entities which was reduced from the 16,087 entities that were detected initially seen in the sample summary. The new entities list was exported as a CEF file using the Export for Recursion function which was then loaded into Profinder under a targeted workflow and only these entities were inspected and manual integrations performed where necessary.

The new data generated in Profinder was then input to MPP and a new Venn diagram constructed which due to better peak integration accomplished by increased data interrogation the number of entities only present in each species was reduced to 7 in beef, 22 in chicken, 1 in fish and 3 in pork (**Fig. 6**). There were 8 entities present in all samples which were used to find a reference compound to confirm that the sample contains gelatin. A CEF file for these 8 entities was

used to extract the data files in Profinder and a compound with similar intensity and peak shape across all of the samples was chosen. This compound had a relative standard deviation across all of the samples in the analysis of 11.84%, a mean peak area of 4779836.6 and a mass neutral mass of 840.4471 found in both the doubly and singularly charged form. The compound eluted at 18.82 min. The reference compound peak was lower in intensity than any of the species identifying peaks in order for species identification to be made to confirm gelatin was present.

The LC-MS conditions and sample preparations used for the initial analysis were used for the MS/MS analysis. The best ion was selected for each compound and a target MS/MS run with collision energies of 10, 20 and 40V and a retention time window of 1 min.

The *porcine* identifying compound found by MPP corresponded to a neutral mass of 1375.6478 and was a doubly charged ion with two attached protons. Mass 1375.6478 fragmented well at all collision energies used. The pork identifying compound found in MPP had the most abundant sodiated peak with an m/z of 805.3806, and a neutral mass of 782.3908, which did not fragment in a well-defined manner. However, there was also a doubly charged ion which had two attached protons at m/z 392.2027 which gave good fragmentation with a CID of 10V and 20V but was heavily fragmented at 40V. The identifying compound which was a singly charged protonated ion with an m/z of 810.4342, and a neutral mass 809.4269, fragmented well with a CID of 40V.

The *bovine* identifying compound found in MPP had neutral mass 756.4011 and was a singly charged ion with one attached proton with m/z of 757.4084 which gave the best fragmentation at 40V. The most abundant doubly charged beef identifying compound with m/z of 428.2028 and a neutral mass 854.3910 gave good fragmentation at 10 and 20V but was heavily fragmented at 40V. The singly charged m/z of 430.2661 and neutral mass 429.2588 gave good fragmentation at 10 and 20V but was also heavily fragmented at 40V.

The *chicken* identifying compound had a neutral mass 881.4235 and was a singly charged ion with one attached proton with m/z of 882.4308, which gave the best fragmentation using a CID of 40V. That with an m/z of 399.2601 was singly charged and a neutral mass 398.2528 gave good fragmentation at 10 and 20V but was heavily fragmented at 40V. Finally the singly charged m/z of 527.3184 and neutral mass 526.3111 gave good fragmentation at 10 and 20V but was heavily fragmented at 40V.

The *fish* identifying compound was a neutral mass 1903.9013 and was a doubly charged ion with two attached protons with m/z of 952.9585, and gave the best fragmentation using a CID of 40V.

3.4 Identification of biomarker peptide sequences

The MS/MS data was searched using Mascot software against the Uniprot sequence database but no matches were found using both trypsin and semi-trypsin digest options. The specificity of trypsin is largely determined by the negatively charged aspartate residue (Asp 189) located in the catalytic pocket (S1) which binds positively charged lysine and arginine residues. Trypsin is an endopeptidase predominantly cleaving proteins within the polypeptide chain at the C-terminal side of lysine and arginine residues. The amino-acid sequence for the α-1 chain of porcine type 1 collagen was aligned with those for beef, chicken and fish. As trypsin was used to digest the gelatin after each arginine (R) or lysine (K) the protein is expected to be cleaved to give the peptides analysed. If these peptides show a different amino acid sequence between the species it could be an identifying peptide for that species against the others. The mono isotopic mass and the m/z of the 2+ or the 3+ ions were then calculated for the differentiating peptides and these targets were searched for in the data for the different gelatin types. With trypsin digested samples it is expected that the ions would predominately be in the 2+ charge state due to the positively charged side chain of lysine or arginine and the positive charge on the N-terminal amino group; any negative charges on carboxylates being protonated in the mass spectrometer ion source. A 3+ charge may occur, for example, due to any histidine side chains.

When the amino acid sequence for porcine type 1 collagen (Q6H2X9) was aligned with bovine type 1 collagen (Q95ND8) and cleaved at predicted trypsin digestion points there were 3 possible peptide fragments predicted (**Table 1**) that had different amino acid sequences between the different species. When aligned with chicken type 1 collagen there were 5 possible peptide fragments predicted that differed in the amino acid sequence (**Table 2**) and when aligned with fish type 1 collagen there were 6 possible peptides (**Table 3**) differing in amino acid sequence.

Although these Tables predict that trypsin digests of gelatin peptides derived from porcine type 1 collagen have some amino acid sequences which are different to those from bovine, chicken or fish type 1 collagen when these peptides are searched against the raw data using the most likely m/z they were not found in the samples prepared in this work. This could be due to different types of collagen being used to manufacture the gelatin or that these peptides were hydrolysed during the gelatin manufacture process. The differentiating peptides found by MPP may be hydrolysed fragments of the predicted peptides or they may be a result of post transcriptional modifications.

3.5 Peptide identification using PEAKS software and Uniprot database

The MS/MS spectra generated for identifying the species specific peptides were analysed using the *de-novo* sequencing tool in PEAKS software with the peptide sequence having the best average local confidence (ALC) score being

reported. Where more than one peptide had the best ALC score they were both reported. The *de-novo* sequence was then searched using the BLAST search function on the UniProt database.

The predicted best match sequence for the *pork* identifying peptide with neutral mass 1375.6478 was AGPAGPDGPLGPAGSR with a local confidence score of 91%. This peptide has a 2+ charge and terminates with an arginine residue as expected from a trypsin digested peptide. This peptide aligned in the BLAST search with an uncharacterised protein, not necessarily from collagen, but present in pigs. The predicted best match sequence for neutral mass 782.3908 was LNGPAPGR with a local confidence score of 82%. This peptide also has a 2+ charge and terminates with an arginine. The predicted best match sequence for neutral mass 809.4269 was LGPLGSPGL with a local confidence score of 96%. This peptide only has a 1+ charge and does not contain either arginine or lysine and is not a typical tryptic peptide.

The predicted best match sequence for the *beef* identifying peptide of neutral mass 756.4011 was WVGLLGL with a ALC score of 90%. This peptide only has a 1+ charge and does not contain either arginine or lysine and so is not a typical tryptic peptide. The predicted best match sequence for neutral mass 854.3910 was WGGPEGPR with a local confidence score of 95%. This peptide has a 2+ charge and terminates with an arginine. There are two best match sequences for the peptide with neutral mass 429.2588 GLAGL and LGAGL and both have a ALC score of 92% and are singly charged peptides not containing arginine or lysine.

The predicted best match sequence for the *chicken* identifying peptide with neutral mass 398.2528 is LGPL with a local confidence score of 98%. This peptide only has a 1+ charge and does not contain either arginine or lysine. The predicted best match sequence for neutral mass 526.3111 is QLGPL with a local confidence score of 95%. This peptide only has a 1+ charge and also does not contain either arginine or lysine. The predicted best match sequence for neutral mass 881.4235 is GALGPLGAVGA with a local confidence score of 82% and does not contain either arginine or lysine.

The predicted best match sequence for the *fish* identifying petide with neutral mass 1903.9013 is AGPLGPTGPAGWLDLGGLQQ with a local confidence score of 83%. This peptide has a 2+ charge but does not contain either arginine or lysine. The 14^{th} amino acid was not glycine as is thought to be the case with collagen peptide sequences.This peptide aligned in the BLAST search with a protein from the bacterium *Burkholderia ambifaria*.

In summary, for each identifying peptide found by MPP it was possible to predict a sequence with a good local confidence score using the PEAKS de-novo sequencing tool. This was not possible with Mascot software even though two of the sequences found by PEAKS, AGPAGPDGPLGPAGSR, (pork) and AGPLGPTGPAGWLDLGGLQQ

(*Burkholderia ambifaria)* were in the UniProt database. The *Burkholderia ambifaria* peptide found in all of the fish gelatin samples tested sugests that some or all of the fish used to make the gelatin were infected with *Burkholderia ambifaria* which are a group of closely related bacteria (Vandamme and Dawyndt, 2011). The predicted pork identifying peptides did not match any of the PEAKS sequenced peptides even if a non perfect match was accepted.

3.6 Analysis of gelatin prepared under controlled conditions

One of the main difficulties encountered was to find a reliable source of uncontaminated gelatin which had fully traceable origins. In order to confirm that the pork, beef, chicken and fish gelatin that the prediction model was based on and that low level traces of different species peptides are from contamination, pure species specific gelatin samples were prepared.

For each species of gelatin all of the predictive peaks were seen (**Table 4**) and were confirmed by accurate mass, retention time and MS/MS fragmentation patterns. There was no incorrect peptide found in the wrong sample type. These laboratory prepared gelatin samples had very similar TIC's to the commercial samples but the concentrations varied due to them not being dried prior to trypsin digestion.

It was concluded that the marker peptides which were found using MPP are also found in the laboratory prepared species specific gelatin samples.

3.7 Identification of 21 unknown gelatin samples using the peptide identifiers

21 samples of unknown gelatin were supplied by Healan ingredients and analysed by LC-MS in a blind test. The peptide identifiers were then searched for in the MS data and identifications made. All 21 unknown samples and one hydrolysed beef collagen (HBC) sample were analysed along with a known sample of pork, beef, chicken and fish gelatin using the same method as was used to find the identifying peptides described earlier.

The results are shown in **Table 5**. The species origin of all 21 samples was predicted correctly. Although in some cases a small peak for a different species was seen this was always smaller than the peak used to characterise gelatin (mass 840.4471) and was ignored. Sample 11 was chicken gelatin but only had one intense peak (526.3111) with small peaks for the other two 881.4235 and 398.2528. This sample also had small peaks for all of the pork peptides. Sample 19 was a pork gelatin but the peak for 809.4269 was not seen but strong peaks were seen for 1375.6478 and 782.3908. Many of the samples were also blends of different batches which were said to be from the same species**.**

The species of all of the samples were correctly predicted when only high intensity peaks were used. The small peaks seen in some of the samples could be caused by carry over from a previous chromatographic run or contamination either at sampling or in manufacture. In this analysis some samples had 'missing' or small peaks, where they were predicted to be present with a strong signal intensity. This could have been due to these peptides being hydrolysed further during gelatin production giving rise to smaller peaks when analysed. None of the samples showed strong peaks in the wrong sample set meaning if a strong peak was seen that could be used to predict the species without the confirmation from the other peptides.

4. Conclusion

No differentiation in the species of origin of gelatin could be obtained by solvent extraction.

Trypsin digestion however showed a good species separation but gave a large spread in the data for each species of gelatin. Using the Profinder and MPP software there were 3 peptides uniquely found only in pork gelatin, 7 only in beef gelatin, and 22 only in chicken gelatin and only 1 only in fish gelatin. *De-novo* sequencing using Peaks software was able to predict matching sequences for the identifying peptides. There was only one peptide found in all of the fish gelatin samples tested and this appears to be from a bacterium. There is a large species diversity in fish gelatin with different species having different characteristics (Jamilah and Harvlnder, 2001) so finding a marker peptide for fish gelatin is difficult indicating that a proteomics aproach to detecting fish gelatin may not be suitable. Peptide 1903 can only be used to detect the presence of protein B1FNW6 from the bacteria *Burkholderia ambifaria* which apears to be commonly found in all fish.

The method developed can beused to identify the species of origin of commercial gelatin samples. Gelatin synthesised under controlled laboratory conditions not only showed the presence of all of the marker peptides but also did not have any of the peaks for other species. The untargeted workflow used here could also be useful to differentiate gelatin from other species that may not have had their collagen sequenced.

Acknowledgements

SW thanks Agilent Technologies and IPOS for financial support.

The authors declare no competing financial interest and any conflict of interest.

References

Alves, P., Arnold, R.J., Clemmer, D.E., Li, Y., Reilly, J.P., Sheng, Q., Tang, H., Xun, Z., Zeng, R., Radivojac. P., 2008. Fast and accurate identification of semi-tryptic peptides in shotgun proteomics. Bioinformatics. 24,102-9.

Anderson, L. N., Anderson, N. G. 1998. Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, 19, 1853-1861.

Brodsky, B., Ramshaw, J. A., 1997. The collagen triple-helix structure. *Matrix Biol.*, 16, 545-554.

Cottrell, J. S. 2011. Protein identification using MS/MS data. *J. Proteomics*, 74, 1842-1851 doi:10.1016/j.jprot.2011.05.014

Darie C. C., *Mod. Chem. appl.* 2013 1:e105. doi:10.4172/2329-6798.1000e105.

Jamilah, B., Harvlnder, G. K. *Fish gelatin from tilapia skin*. 2001, Putra: Universiti Putra Malaysia.

Liu, H., Zhang, J., Hanchang, S., Xu, C., Zhu, Y., Xie, H. 2011. The Prediction of Peptide Charge States for Electrospray. *Proc. Environ. Sci.* 8, 483-491.

Krenkova, J., Lacher, N. A., Svec, F. 2009. Highly efficient enzyme reactors containing trypsin and endoproteinase LysC immobilezed on porous polymer monolith coupled to MS suitable for analysis of antibodies. *Anal. Chem.*, 81, 2004-2012.

Mass Profiler Professional Software. (2017, January 19). Retrieved from Agilent Technologies: http://www.agilent.com/en-us/products/software-informatics/masshunter-suite/masshunter-for-life-science-research/mass-profiler-professional-software

Maynes, R., 1987. Ed. *Structure and function of collagen types.* 1st ed Elsevier

Medzihradszky, K. F., Chalkley, R. J. 2013. Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spec. Rev.*, 34, 43-63.

Mitchell, P., 2010. Proteomics retrenches. *Nature Biotechnology*, 28, 665-670.

Naegele, E. 2011. *Making your LC Method Compatible with Mass Spectrometry*, Agilent Technologies. https://www.agilent.com/cs/library/technicaloverviews/public/5990--7413EN.pdf

Olsen, J. V., Ong, S.-E., Mann, M. 2004. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Molecular and Cellular Proteomics*, 3, 608-614.

Orton, D. J., Doucette, A. A., 2013**,** Proteomic Workflows for Biomarker Identification Using Mass Spectrometry — Technical and Statistical Considerations during Initial Discovery, *Proteomes*, 1, 109-127

Phillips, G. O., Williams, P. A., 2009. *Handbook of hydrocolloids.* CRC Press.

Rodriguez, J., Gupta, N., Smith, R. D., Pevzner, P. A., 2008. Does Trypsin Cut Before Proline? *J. Proteome Res.*, 7, 300-305.

Roepstorff, P., Fohlman, J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*, 11, 601.

Sana, T. (2017, January 19). *Profinder webinar.* Retrieved from Agilent library: http://cn.agilent.com/cs/library/eseminars/public/Profinder%20Webinar_Final_17Dec13.pd

Steen, H., Mann, M., 2004. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.*, **5**, 699-711.

Vandamme, P., Dawyndt, P. 2011. Classification and identification of the *Burkholderia cepacia* complex: Past, present and future. *Syst. Appl. Microbiol.*, 34, 87-95.

Venien, A., Levieux, D., 2005. Differentiation of bovine from porcine gelatins using polyclonal anti-peptide antibodies in indirect and competitive indirect ELISA. *J. Pharm. Biomed. Anal.*, 37, 418-424.

Yang, C. T., Ghosh, D., Beaudry, F., 2018 Detection of gelatin adulteration using bio-informatics, proteomics and high-resolution mass spectrometry ***Food Additives & Contaminants: Part A,*** doi: 10.1080/19440049.2017.1416680

Zhang, G., Liu, T., Wang, Q., Chen, L., Lei, J., Luo, J., Su, Z., 2009. Mass spectrometric detection of marker peptides in tryptic digests of gelatin: Anew method to differentiate between bovine and porcine gelatin. *Food Hydrocolloids*, 23, 2001-2007.