Malaysian Applied Biology (2024) 53(3): 255-266
https://doi.org/10.55230/mabjournal.v53i3.2806

Research

Bioinformatics Tools Assist in The Screening of Potential Porcine-Specific Peptide Biomarkers of Gelatin and Collagen For Halal Authentication

Muhd Syarafuddin Abdul Shukor[1], Mohd Faiz Foong Abdullah[2], Siti Aimi Sarah Zainal Abidin[1,3]*, Mohd Hafis Yuswan[4], Azilawati Mohd Ismail[5]

1. Food Science and Technology Program, Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
2. Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
3. Malaysia Institute of Transport (MITRANS), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
4. Laboratory of Halal Services, Halal Products Research Institute, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
5. Food Safety and Quality Laboratory, Selangor State Health Department, Ministry of Health, 41200 Klang, Selangor, Malaysia

*Corresponding author: sitiaimi@uitm.edu.my

ABSTRACT

Gelatin and collagen are two animal-derived ingredients that are widely used in various industries. Both have distinctive physico-chemical characteristic that made them ingredients of interest for many industrial players to be applied as there are vast arrays of usage in the food, cosmetic and biomedical fields. However, the origin of gelatin and collagen poses ethical and religious concerns, especially for Muslims and Jews who have restrictions on food consumption. Porcine by-products are of concern for religious and health reasons, and there is a demand for precise and reliable detection techniques. The limitation of DNA detection is due to extreme environment in food processing which results in low extractability of DNA. Therefore, peptide-based detection using mass spectrometry is required. However, identify the suitable marker is like searching needle in haystacks. Hence, combination of bioinformatics and mass spectrometry is proposed. This study aims to identify the specific peptide biomarkers by employing bioinformatics technique which can be applied to identify gelatin and collagen sources with the aid of mass spectrometry. In these approach, combination of Petunia Trans-Proteomic Pipeline (TPP, version 5.2.0) and sequence alignment ClustalW were applied to facilitate the MS data (LC-QTOF-MS) and peptide identification. As a result, 69 fasta file of protein sequence from both UniProtKB and NCBInr have been collected, 81 collagen peptides sequence and 118 gelatine peptides has been attainable that have the potential to distinguish different species. In conclusion, in silico protein sequence approaches helps to enable rapid screening of proteotypic peptides that can serve as species biomarkers proficiently.

Key words: Bioinformatics, collagen, gelatin, peptide, porcine

Article History
Accepted: 19 July 2024
First version online: 30 September 2024

Cite This Article:
Shukor, M.S.A., Abdullah, M.F.F., Abidin, S.A.S.Z., Yuswan, M.H. & Ismail, A.M. 2024. Bioinformatics tools assist in the screening of potential porcine-specific peptide biomarkers of gelatin and collagen for halal authentication. Malaysian Applied Biology, 53(3): 255-266. https://doi.org/10.55230/mabjournal.v53i3.2806

INTRODUCTION

Gelatin is an animal-derived food ingredient which is commonly used in various industries and consist of a mixture of polypeptides that originated from the fibrous protein called collagen. During the production of gelatin, collagen will undergo a partial hydrolysis process where dilute alkali or acid is used to accomplish partial cleavage of crosslink thus allowing the structure to break and form “warm-water soluble collagen” specifically known as gelatin (Yilmaz *et al*., 2013). Gelatin is primarily extracted from the skin of animals containing collagen type I and type III, cartilages with collagen type II and bones comprising collagen type I and type II which are mainly pork followed by beef and fish (Yang *et al*., 2018). According to Boran and Regenstein (2010). Approximately, 80% of worldwide gelatin production is derived from pork or porcine derivatives. Among mammalian gelatins, porcine gelatin is most commonly utilized in various industrial

applications (Cebi *et al*., 2019). The approximate amino acid composition of gelatin is as follows: glycine 21%, proline 12%, hydroxyproline 12%, glutamic acid 10%, alanine 9%, arginine 8%, aspartic acid 6%, lysine 4%, serine 4%, leucine 3%, valine 2%, phenylalanine 2%, threonine 2%, isoleucine 1%, hydroxylysine 1%, methionine and histidine < 1%, and tyrosine < 0.5% (Karidu raganavar *et al*., 2014).

The unique structure of gelatin mimics the structure of its parent collagen allowing it to have elasticity, softness and reversible gel properties (Yilmaz *et al*., 2013). Moreover, other unique physicochemical properties of gelatin have wider applications in a diverse range of fields. For example, porcine gelatin is incorporated in a variety of chewable and gummy sweets. Besides the food industry, pharmaceutical companies utilize gelatin in some of their products such as soft gel and mini-capsules. Gelatin, a hydrolyzed derivative of collagen, serves as a fundamental and vital structural protein. It constructs flexible and resilient molecular networks that subsequently extend the tendons and elastic layers. This mechanism provides indispensable and essential support for the skin, the integumentary system, and various internal organs across a multitude of diverse species. (Silvipriya *et al*., 2015). Structurally, collagen is a protein with a triple helical structure of three polypeptide chains (Buckley, 2016). Each chain is composed of repeating amino acids which are Gly-X-Y where the X and Y positions are commonly constituted of proline and hydroxyproline respectively (Hashim *et al*., 2010; Silvipriya *et al*., 2015). The sources of collagen mostly from bone and cartilage of bovine and porcine and there were nearly 28 types of collagens that have been reported in previous studies (Hashim *et al*., 2010; Silva *et al*., 2014; Silvipriya *et al*., 2015). According to Hashim 2010, the most common collagens found are types I, II, III, IV and V. Like gelatin, collagen also has numerous applications in the food, cosmetic, biomedical, and pharmaceutical industries. the collagen-based products market is expected to reach $293 million in 2020, especially for collagen supplements, which is well above $50 million in 2014.

Nevertheless, the animal origin of gelatin and collagen used in various industries is of concern to certain sections of consumers. The use of porcine as a consequence of low cost and easy accessibility had caused much concern among Muslims and Jews due to religious restrictions on food and diet. These days, the rise of economic needs among industries has led to the internationalization of food production, distribution and retailing (Ballin *et al*., 2009; Manning & Soon, 2016; Stadler *et al*., 2016). These can induce food industries and mass-market product manufacturers to source cheaper and prohibited ingredients as a substitution in prioritizing profit. Hence, it elevates the adulteration or deceptive blending of animal products or by-products, which can vary from misleading label assertions to the incorporation of additives (Yang *et al*., 2018). In 2022, the Brunei Minister of Religious Affairs (MoRA) confirmed that White Rabbit milk candy contains pig gelatin in its ingredients after a laboratory test was conducted. It is further supported by a series of laboratory tests conducted by the Islamic Development Department of Malaysia (JAKIM) (Husin, 2022). In the same year, similar cases arise involving pork gelatin in food. A product from Assorted Chewy sweets called Fruit-tella was confirmed to use pork gelatin after thorough inspections by the Halal Food Control Division and has been confirmed by the manufacturer through the official website. Back in 2020, a major crime involving an illegal meat cartel has been reported in Malaysia. The recent case had caused much damage to Malaysia's reputation as an international halal hub. As much as 1,500 tons of meat have been found in the arrest of illegal meat cartel smuggling syndicates from four foreign countries. Indeed, with the widespread food fraud cases, authenticity and transparency in the food industry are vital to cease the concern about halal authenticity and fraud.

To date, identifying porcine by-products like gelatin and collagen has been a subject of interest among academics and researchers worldwide. However, DNA based detection method has a serious limitation when it comes to contamination, whereby it happens at any stage from food production, and preparation as well as transportation of products. In a previous investigation, it was revealed that during gelatin processing, DNA integrity is destroyed due to elevated temperature and acidic conditions used in the manufacturing process (Jaswir *et al*., 2016; Sha *et al*., 2018). Another alternative method is by immunological approach namely ELISA. Although the method is simple, cheap and suitable for routine analysis, It can only detect a restricted sequence of the proteins (i.e., epitope), the precision of the detection is entirely reliant on the specificity of the antibody employed, hence it poses a considerable risk of false negatives or positives and it does not allow for any data mining or derivation of molecular information (Tighe *et al*., 2015).

Meanwhile, one promising method in species identification is by proteomics-based technique. A step has been taken to implement the proteomics advances in meat science research, and recently in food science and nutrition, or known as foodomics. The term foodomics refers to the study of the application of advanced omics techniques (Cifuentes, 2017), including transcriptomics, metabolomics, and proteomics. This field encompasses aspects of food quality, safety, and the discovery of new

bioactive components within food as studies of the application of advanced omics approaches to food (Jagadeesh *et al*., 2017). These days, the trend among society that is deemed for transparency in food authentication has risen as a result of religious concerns and health (Meijer *et al*., 2021). In assessing this matter, proteomics plays a major role for quantification, adulteration detection and identification of species-specific protein biomarkers in food with the aid of mass spectrometry methods (Afzaal *et al*., 2021). Recently, the field of food science has seen an increased application of proteome research which aids in the quality and safety of fish and the identification of species for protein-rich food products such as meat, fish, crustaceans and milk. Proteome research also plays a crucial role in detecting undeclared substances from plant material and even in identifying markers related to the geographic origin of certain species (Calvano *et al*., 2013; Robin Korte & Brockmeyer, 2017). To add up, proteomic approaches in food offer rapid identification, and comprehensive and provide in-depth analysis down to the peptide level (Gallardo *et al*., 2013). This proves that proteomics can meet the consumer's demand regarding food authentication and mislabeling issues (Meijer *et al*., 2021). It was said that the LC-MS/MS approach offers more accurate and reliable gelatin speciation than other methods (Grundy *et al*., 2016). In the proteomic workflow, analysis of proteome normally will be conducted after the preparation of the sample by either one or separation steps and followed by mass spectrometric analysis (Ortea *et al*., 2016).

Therefore, bioinformatic analysis will be integrated with mass spectrometry identification to ease the species-specific screening process. As sample complexity increases, high selectivity, sensitivity and wide dynamic range of instruments are needed. This is when the MS system comes in handy when performing such complex assay (Orduna *et al*., 2015). It is an analytical method used to detect and measure sample analytes in gaseous form based on their mass-to-charge ratio (m/z). In conjunction with the above approaches, bioinformatics tools or computational tools were applied to determine the network of protein interaction, depict the three-dimensional structure of protein and decipher the amino acid sequence code for species identification (Nugraha *et al*., 2022). Previous research conducted by Yang *et al*., (2018), bioinformatics analysis was applied to aligned protein sequences and a systematic species comparison was performed. Nevertheless, there is little information about bioinformatic analysis on the identification of porcine gelatin and collagen. Previous study by Guo *et al*., (2018), peptide identifications of gelatin were thoroughly discussed but only with the means of mass spectrometry approach. In another study comprised of gelatin, the bioinformatic strategy was used for the determination of origin species however, the study scope was only focused on one scope which is gelatin not collagen. Therefore, the present study will emphasize an in silico experiment to identify possible peptide biomarkers for both collagen and gelatin which must be specifically owned by the porcine family to shorten the identification protocol in mass spectrometry.

MATERIALS AND METHODS

Materials

Porcine gelatin from skin and collagen powder were obtained from Sigma-Aldrich (USA) and Bulk supplement (USA). Proteomic grade trypsin with sequencing grade from Promega identification number #V5280 (Madison, USA), Acetonitrile (LC-MS), Ammonium Bicarbonate (NH_4HCO_3), Ultrapure water, Iodoacetamide (IAA), Dithiothreitol (DTT) Formic acid from Thermo Fisher Scientific (USA) and OMIX C18 tip were purchased from Agilent (Santa Clara, USA).

Method of analysis

Five miligram of gelatin and collagen standard were weighed before adding 5 mL of 50 mM of Ammonium Bicarbonate (Zhang *et al*., 2019). The mixture then was vortex for 10 sec and sonicated for 15 min until fully dissolved. 200 µL of samples were then transferred into a centrifuge tube and added with trypsin as follows: 1:20 ratio proteomic-grade trypsin to protein content were added by the manufacturer’s instructions as stated in COA (Promega Trypsin Gold, Mass Spectrometry Grade, Part No V5280). Further, the mixture was incubated at 37°C overnight. Subsequently, after being digested, protein samples were alkylated and reduced with the addition of 0.5 mM of DTT and 0.5 mM of IAA in a 1:20 ratio (v/v) at 37°C for 6 hr (Yuswan *et al*., 2021). Afterwards, 2 µL formic acid was added to the mixture to quench the proteolytical digestion (Jannat *et al*., 2020). Thereafter, digested protein samples were desalted using Agilent Bond Elut OMIX C18 tip by protocols described by the manufacturer before injection.

LC-MS/MS mass spectrometry condition

In this study, the chromatographic separation of proteins was conducted using a high-performance liquid chromatography system (Agilent 1200 Series from Waldbronn, Germany). This was supplemented with a Liquid chromatography-electrospray ionisation-quadrupole-time of flight system (LC-ESI-Q-TOF; Agilent 6520, Agilent Technologies, Santa Clara, CA) for additional analysis of the separated samples. The column utilized for the chromatographic separation was a reversed-phase Phenomenex Kinetex Core-shell C18 column with dimensions of 100 × 2.1 mm, particle size of 2.6 µm and pore size of 100 Å. The solvent system employed for separation was a gradient elution with 0.1% of FA in ultrapure water (eluent A) and 100% of can (eluent B). Separation was achieved through gradient elution with 0.1% of FA in ultrapure water (eluent A) and 100% of can (eluent B). Initially, eluent B was set at 3% for 1 min, followed by 97% for 20 min and finally held for 1 min before setting it to the initial condition for 3 min. (Yuswan,2021). Further, the chromatographic separation was performed in triplicate for each sample through three 20-µL injections per injection (n=45). The total run time for the analysis was set at 25 min with a flow rate of 300 µL/min. The detection of the mass spectrum was carried out using electrospray ionization (ESI). Nitrogen gas was used as the drying gas at a temperature setting of 350°C and a flow rate setting of 5.0 L/min. The nebuliser voltage and fragmentor were maintained at settings of 35 psi and 125V respectively. Nitrogen gas was also used for collision-induced dissociation of precursor ions. MS and MS/MS analyses were acquired at m/z range from 50 to 1000 in positive mode.

Data analysis

The resulting mass chromatogram was then transformed into the centroided peaklist files (mzML) by employing a software called ProteoWizard MSConvertGUI (64-bit, Version: 3.0.1908-43e675997) (Palo Alto, USA). The mzML format, an open-source solution, serves as a storage medium for raw data output from mass spectrometers. The generation of the peaklist was facilitated by selecting the filter type of peak picking, with the algorithm parameter set to vendor (MS level = 1-). Subsequently, all mzML files were subjected to an open-source sequence database search engine, Comet (version 2018.01 rev. 4). To execute Comet, the Petunia Trans-Proteomic Pipeline (TPP, version 5.2.0) was employed. An in-house target-decoy database was constructed from a reference proteome of pig (Sus scrofa), identified by UP000008227 and downloaded from UniProtKB. This database encompasses both SwissProt (1436 entries) and TrEMBL (48,357 entries), along with entries from the NCBInr database for protein identification.

For the open-source database search engine, standard modifications were applied. These included setting the precursor and fragment mass tolerance to 0.5 and 0.7 Da, respectively. The cutoff of false discovery rate (FDR) for identification was set at ≤ 1%. Trypsin digestion was configured to be semi-digested with a maximum of one missed cleavage (Amir *et al*., 2021). Variable modifications included carbamidomethylation of cysteine, oxidation of methionine and reduction of asparagine or glutamine residues. For the determination of gelatin protein, therefore, the following modifications were subjected to collagen database due to the inexistence of sequence database: oxidation (P, K, M), and deamidation (N, Q) were set as variable modifications and carbomidomethyl (C) as fixed modification. Additional parameters included peptide tolerance ± 25 ppm, MS/MS tolerance ± 0.5 Da, 2 missed cleavage and an ion score >40 (Sha *et al*., 2019; Zhang *et al*., 2014). Further, the generated pepXML files were then subjected to Peptide Prophet and Protein Prophet to further analyse as referred to in Figure 1, the flow of TPP.

Identification of peptide biomarkers

Further, multiple sequence alignments using ClustalW were conducted. The resulting list of proteins identified from the MS/MS profiling was then analyzed using multiple sequence alignment to ascertain the unique amino acid site of a specific protein in sus scrofa. The FASTA sequences of all the detected proteins were subsequently obtained from the NCBInr protein database (http://www.ncbi.nlm. nih. gov) and UniprotKb (http://www.uniprot.org/) which include accession numbers, names and species. Subsequently, subjected to sequence homology or similarity analysis using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). To acquire optimal alignment results, the following parameters were implemented which were commonly used as default for global alignment: Matrix: Gonnet, GAP OPEN: 10, GAP EXTENSION: 0.2, GAP DISTANCES: 5, NO END GAPS: no, ITERATION: none, NUMBER: 1, CLUSTERING: NJ, OUTPUT FORMAT: clustalw/ numbers, and ORDER: aligned (Sarah *et al*.,2016).

Election of peptide biomarkers

The selected specific peptide is denoted as a “proteotypic peptide” where there’s a selection of

criteria to follow including possessing a minimum sequence length of peptide that is allowed is 6 and only having a probability between 0.95 – 1 indicates a high percentage of occurrence and a distinct sequence that is only seen in porcine.

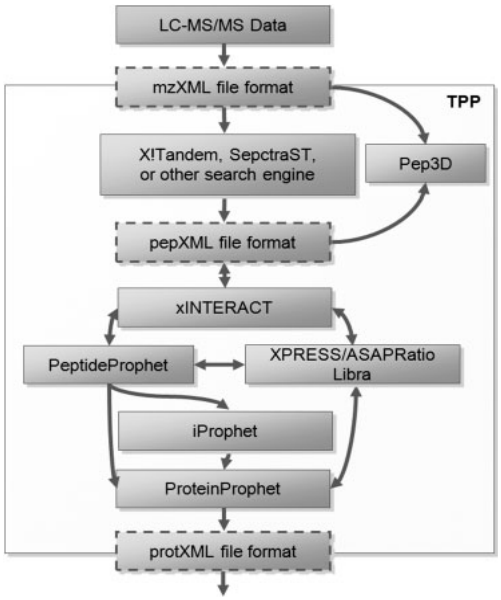

Fig.1. Schematic overview of Trans-Proteomic Pipeline workflow (Deutsch *et al*., 2010)

RESULTS AND DISCUSSION

Sequence curation

Collagen protein sequences derived from porcine, bovine, fish and chicken were accumulated from the Universal Protein (UniProt) and NCBI. In total of 69 fasta files of all the sequences of collagen were summarized in Table 1.

Table 1. Sequences for identification of collagen biomarker peptides

Number	Accession Number	Protein Type	Origin Organism	Length
1	>tr|A0A286ZQ85|A0A286ZQ85	Collagen type III alpha 1 chain	Sus Scrofa (Porcine)	1467
2	>tr|F1SFA7|F1SFA7	Collagen type I alpha 2 chain	Sus Scrofa (Porcine)	1365
3	tr|A0A5G2RJ53|A0A5G2RJ53	Collagen type IV alpha 2 chain	Sus Scrofa (Porcine)	1663
4	tr|A0A287BLD2|A0A287BLD2	Collagen type I alpha 1 chain	Sus Scrofa (Porcine)	1348
5	tr|A0A287BPM1|A0A287BPM1	Collagen type V alpha 2 chain	Sus Scrofa (Porcine)	1474
6	tr|A0A286ZWS8|A0A286ZWS8	Collagen type II alpha 1 chain	Sus Scrofa (Porcine)	1417
7	tr|F1S021|F1S021	Collagen type V alpha 1 chain	Sus Scrofa (Porcine)	1918
8	tr|A0A5G2QHA4|A0A5G2QHA4	Collagen type XII alpha 1 chain	Sus Scrofa (Porcine)	2989
9	tr|A0A5G2R4W0|A0A5G2R4W0	Collagen type XVIII alpha 1 chain	Sus Scrofa (Porcine)	1238
10	tr|A0A287AFT5|A0A287AFT5	Collagen type XV alpha 1	Sus Scrofa (Porcine)	1317
11	>sp|P02453|	Collagen alpha-1(I)	Bos Taurus (Bovine)	1463
12	>sp|P02459| - U	Collagen alpha-1(II)	Bos Taurus (Bovine)	1487

Table 1. Continued

Number	Accession Number	Protein Type	Origin Organism	Length
13	>>AAI23470.1	Collagen type III, alpha 1	Bos Taurus (Bovine)	1466
14	>tr|F1N7Q7|F1N7Q7	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1529
15	>tr|A0A3Q1NK70|A0A3Q1NK70	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1539
16	>tr|A0A3Q1MPS8|A0A3Q1MPS8	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1548
17	>tr|A0A3Q1LZF7|A0A3Q1LZF7	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1546
18	>tr|A0A3Q1LT89|A0A3Q1LT89	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1496
19	>tr|A0A3Q1LLV2|A0A3Q1LLV2	Collagen alpha-2(IV)	Bos Taurus (Bovine)	1551
20	>tr|G3MZI7|G3MZI7	Collagen type V alpha 1	Bos Taurus (Bovine)	1847
21	>tr|K7QCV6|K7QCV6	collagen type V alpha 1 (Fragment)	Bos Taurus (Bovine)	1014
22	>tr|A0A3Q1MDT9|A0A3Q1MDT9	Collagen type V alpha 2	Bos Taurus (Bovine)	1336
23	>tr|A0A3Q1LL95|A0A3Q1LL95	Collagen type V alpha 2	Bos Taurus (Bovine)	1240
24	>tr|A0A3Q1LSX2|A0A3Q1LSX2	Collagen type V alpha 2	Bos Taurus (Bovine)	1318
25	>tr|F1N2Y2|F1N2Y2	Collagen type V alpha 2	Bos Taurus (Bovine)	1294
26	>XP_005210721.2	collagen alpha-1(XII) chain isoform X1	Bos Taurus (Bovine)	3115
27	>XP_005210722.2	collagen alpha-1(XII) chain isoform X2	Bos Taurus (Bovine)	3065
28	>XP_010806670.2	collagen alpha-1(XII) chain isoform X3	Bos Taurus (Bovine)	1951
29	>XP_024851190.1	collagen alpha-1(XV) chain isoform X1	Bos Taurus (Bovine)	1376
30	>XP_024851191.1	collagen alpha-1(XV) chain isoform X2	Bos Taurus (Bovine)	1372
31	>XP_024851192.1	collagen alpha-1(XV) chain isoform X3	Bos Taurus (Bovine)	1344
32	>XP_024830782.1	collagen alpha-1(XVIII) chain isoform X1	Bos Taurus (Bovine)	1770
33	>XP_024830811.1	collagen alpha-1(XVIII) chain isoform X2	Bos Taurus (Bovine)	1762
34	>XP_024830844.1	collagen alpha-1(XVIII) chain isoform X3	Bos Taurus (Bovine)	1761
35	>XP_024830882.1	collagen alpha-1(XVIII) chain isoform X4	Bos Taurus (Bovine)	1523
36	>XP_024830923.1	collagen alpha-1(XVIII) chain isoform X5	Bos Taurus (Bovine)	1329
37	>tr|A0A1D5PYU1|A0A1D5PYU1	Collagen alpha-1(I)	Gallus Gallus (Chicken)	1453
38	>sp|P02460|CO2A1	Collagen alpha-1(II) chain (Fragment)	Gallus Gallus (Chicken)	859
39	>tr|A0A5H1ZRK9|A0A5H1ZRK9	Collagen alpha-1(II)	Gallus Gallus (Chicken)	1420
40	>sp|P02460-2|	Collagen alpha-1(II) chain (Fragment)	Gallus Gallus (Chicken)	859

Table 1. Continued

Number	Accession Number	Protein Type	Origin Organism	Length
41	>XP_040559299.1	collagen alpha-1(III) chain isoform X1	Gallus Gallus (Chicken)	1442
42	>tr|A0A3Q2TTC1|A0A3Q2TTC1	Collagen IV	Gallus Gallus (Chicken)	1847
43	>tr|F1P2Q3|F1P2Q3	Collagen IV	Gallus Gallus (Chicken)	1715
44	>tr|A0A3Q2UBM2|A0A3Q2UBM2	Collagen alpha - 2(V)	Gallus Gallus (Chicken)	1496
45	>XP_046784674.1	collagen alpha-1(V) chain isoform X2	Gallus Gallus (Chicken)	1813
46	>XP_046784673.1	collagen alpha-1(V) chain isoform X1	Gallus Gallus (Chicken)	1813
47	>XP_046769120.1	collagen alpha-1(XII) chain isoform X1	Gallus Gallus (Chicken)	3122
48	>XP_046769121.1	collagen alpha-1(XII) chain isoform X2	Gallus Gallus (Chicken)	3080
49	>XP_046769122.1	collagen alpha-1(XII) chain isoform X3	Gallus Gallus (Chicken)	3065
50	>XP_046769123.1	collagen alpha-1(XII) chain isoform X4	Gallus Gallus (Chicken)	3065
51	>XP_046769126.1	collagen alpha-1(XII) chain isoform X5	Gallus Gallus (Chicken)	1960
52	>XP_046768549.1	collagen alpha-1(XV) chain isoform X1	Gallus Gallus (Chicken)	1134
53	>XP_046768550.1	collagen alpha-1(XV) chain isoform X2	Gallus Gallus (Chicken)	1126
54	>XP_046768551.1	collagen alpha-1(XV) chain isoform X3	Gallus Gallus (Chicken)	1124
55	>XP_046768552.1	collagen alpha-1(XV) chain isoform X4	Gallus Gallus (Chicken)	1116
56	>XP_046768553.1	collagen alpha-1(XV) chain isoform X5	Gallus Gallus (Chicken)	1001
57	>XP_046776806.1	collagen alpha-1(XVIII) chain isoform X1	Gallus Gallus (Chicken)	1506
58	>XP_030197640.1	collagen alpha-1(II) chain-like isoform X3	Godus Morhua (Fish)	809
59	>XP_030231027.1	collagen alpha-1(II) chain isoform X1	Godus Morhua (Fish)	1489
60	XP_030231028.1	collagen alpha-1(II) chain isoform X2	Godus Morhua (Fish)	1419
61	>XP_030198694.1	collagen alpha-2(IV) chain isoform X1	Godus Morhua (Fish)	1642
62	>XP_030209191.1	collagen alpha-1(V) chain isoform X1	Godus Morhua (Fish)	2049
63	>XP_030209192.1	collagen alpha-1(V) chain isoform X2	Godus Morhua (Fish)	2049
64	>XP_030209193.1	collagen alpha-1(V) chain isoform X3	Godus Morhua (Fish)	1961
65	>XP_030212283.1	collagen alpha-1(XII) chain isoform X1	Godus Morhua (Fish)	3114
66	>XP_030212285.1	collagen alpha-1(XII) chain isoform X2	Godus Morhua (Fish)	3107
67	>XP_030212287.1	collagen alpha-1(XII) chain isoform X4	Godus Morhua (Fish)	3100
68	>XP_030212288.1	collagen alpha-1(XII) chain isoform	Godus Morhua (Fish)	3075
69	>XP_030212289.1	collagen alpha-1(XII) chain isoform X6	Godus Morhua (Fish)	1941

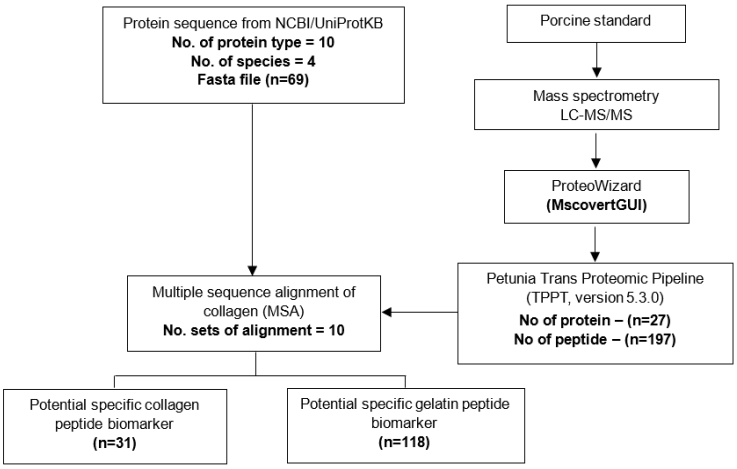

Fig. 2. Flow chart strategy of the analysis steps to indicate peptides sequence of porcine gelatin and collagen

In this study, to determine proteotypic peptides, a customized protein sequence database was generated that allowed accurate porcine gelatin and collagen speciation. As referred to Table 1 above, the 13 types of porcine (sus scrofa) collagen protein that were discovered after analyzing using comet through TPP software were Collagen type 3 alpha 1 chain, Collagen type 1 alpha 2 chain, Collagen type 4 alpha 2 chain, Collagen type 1 alpha 1 chain, Collagen type 5 alpha 2 chain, Collagen type 2 alpha 1 chain, Collagen type 5 alpha 1 chain, Collagen type 12 alpha 1 chain, Collagen type 18 alpha 1 chain, Collagen type 15 alpha 1, Collagen type 16 alpha 1 chain, Collagen type 6 alpha 2 chain and Collagen type 6 alpha 5 chain which can be seen in Figure 3. Nevertheless, three of the collagen protein were then filtered out which is Collagen type 16 alpha 1 chain, Collagen type 6 alpha 2 chain and Collagen type 6 alpha 5 prior to identification of specific peptide marker due to unmet criteria of peptide. With the discovery of the collagen protein, a thorough search through UniProtKb and NCBInr has been made to look for similar types of the 10 porcine (sus scrofa) collagen protein in other organisms namely bovine (bos taurus), chicken (gallus gallus) and fish (godus morhua). Hence, a total of 69 fasta file with accession number of collagen protein composed of four different organisms which involved pig (porcine) bovine (bos taurus), chicken (gallus gallus) and fish (godus morhua) have been compiled.

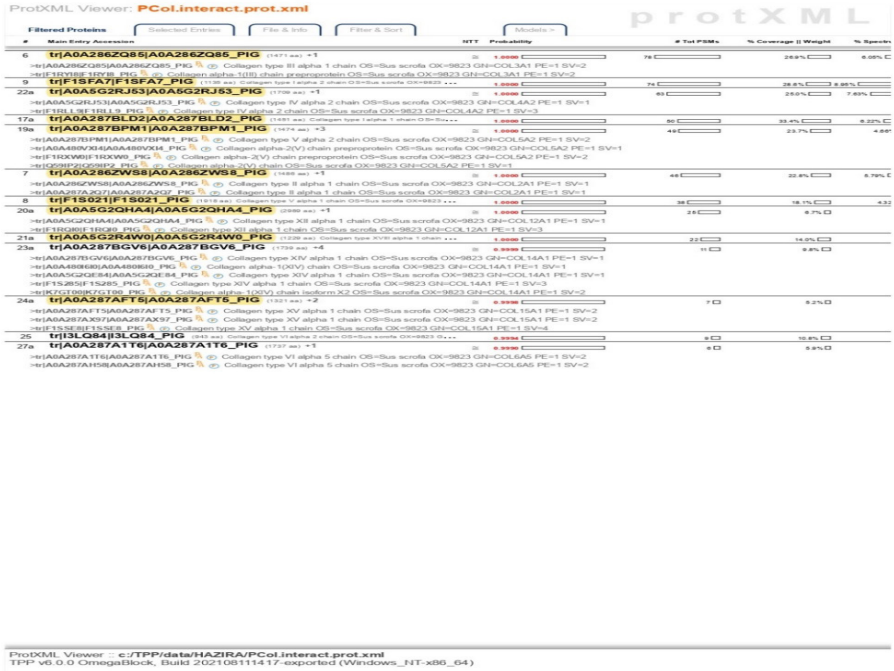

Fig. 3. Collagen protein derived from porcine (Sus scrofa) generated by Petunia Trans-Proteomic Pipeline (TPP, version 5.2.0) after mass spectrometry identification.

Sequence analysis for identification of porcine specific peptide

Continuity from the observed protein, determination of peptide biomarker of species specific were made based on parameter listed above. For gelatin, due to unavailability of gelatin database, no sequence homology was conducted unlike collagen. Hence, a total of 118 specific tryptic peptides from 7 types of collagen proteins that were discovered after analysis was performed directly act as potential specific peptide biomarkers for porcine as can be seen in Table 2. Moving to collagen, in total, there were 81 peptides discovered within the 10 collagen proteins. For collagen type III alpha 1 chain – 13 peptides, collagen type I alpha 2 chain – 16 peptides, collagen type 4 alpha 2 chain – 7 peptides, collagen type 1 alpha 1 chain – 17 peptides, collagen type 5 alpha 2 chain – 6 peptide, collagen type II alpha 1 chain – 7 peptides, collagen type V alpha 1 chain – 9 peptides, collagen type XIII alpha 1 chain – 3 peptides, collagen type XVIII alpha 1 chain – 2 peptides and lastly, collagen XV alpha 1 chain – 1 peptide. Subsequently, a multiple sequence alignment analysis MSA of each identified collagen protein was conducted using CLUSTALW to determine species specific amino acid sequence. All 69 fasta file composed of porcine, bovine, chicken and fish were subjected to sequence homology analysis against porcine sequence. An example of amino acid sequence alignment can be seen in Figure 4. The figure shows that sequence 'GLPGEPGAAWLLGPKGPPG' can be labelled as specific biomarker as no similar sequence was found in other animals.

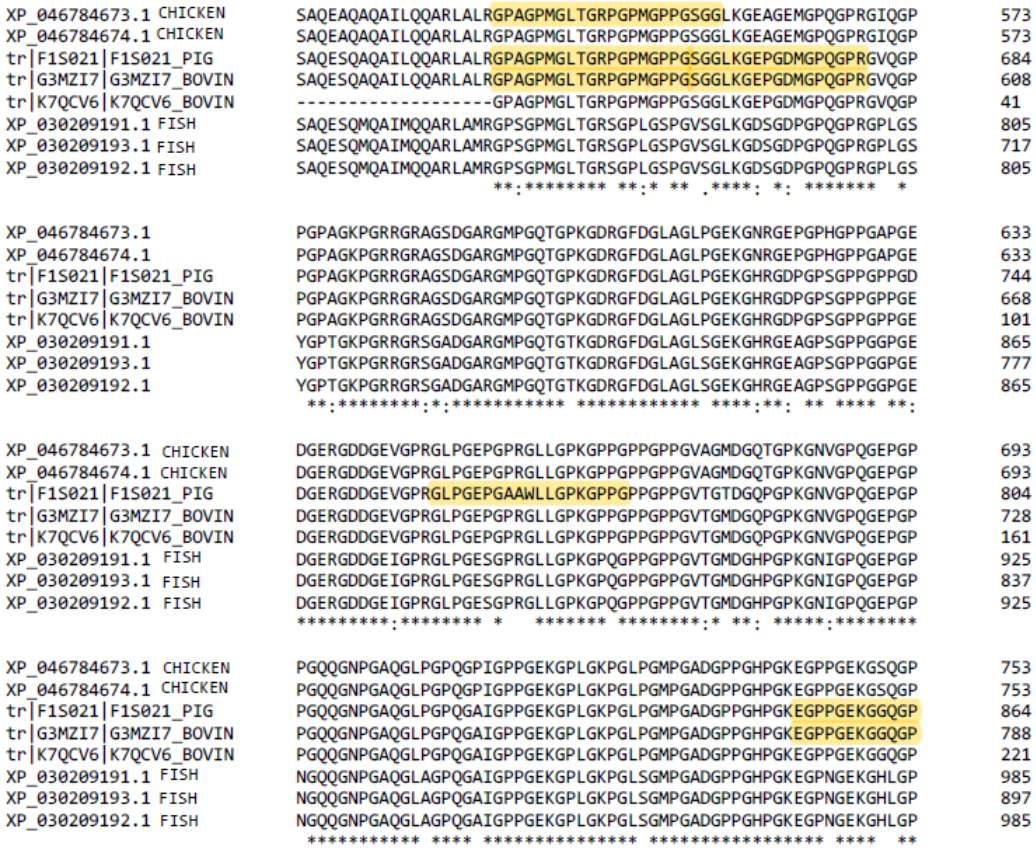

Fig. 4. Multiple sequence alignment of collagen type 5 alpha 1 in four different animals. Accession number: XP_046784673.1 (Chicken), XP_046784674.1 (Chicken), tr|F1S021 (Pig), tr|G3MZI7 (Bovine), tr|K7QCV6 (Bovine), XP_030209191.1 (Fish), XP_030209193.1 (Fish) and XP_030209192.1 (Fish). The sequence alignment was constructed using ClustalW.

In a further stage, the 81 proteotypic peptides of collagen were then tabulated in Supplementary Table 2 after being subjected to MSA analysis. All peptides involved in each protein have been categorized as either its specific peptide biomarker of porcine or belongs to another organism. From the table, it can be seen that two peptides which are 'GPVGPSGPPGK' and 'GPVGPSGPPGKDGASGHPGP' in collagen type 3 alpha 1 chain are not specific. For collagen type 1 alpha 2, out of sixteen peptides, only six that are specific which are 'GERGPPGESGAAGPAGPIG', 'AAGPAGPNGPPGPAGSR', 'VGAAGPAGPNGPPGPAGSR', 'GFPGPKGPTGDPGK', 'APGAVGAPGPAGANGDR' and 'APGAVGAPGPAGANGDR'. Moving on, in the collagen type 4 alpha 2 chain, only one out of seven peptides are not a specific porcine marker which is 'GGRGQPGPVGPQGYTG'. As in

collagen type 1 alpha 1, four specific peptides namely 'GPAGERGSPGPAGPK', 'RGSPGPAGPK', 'GEPGPPGPAGAAGPAGNPGADGQPGG' and 'GEPGPPGPAGAAGPAGNPGADGQPGGKGANGAP' were determined. Next, there are no specific biomarkers that were determined in collagen type 2 alpha 1 as it corresponds to other sequence organisms. For collagen type 5 alpha 1 chain, two peptides from nine are specific which are 'GLPGEPGAAWLLGPKGPPG' and 'GEAGHPGPPGPPGPPGE'. As for collagen type 12 alpha 1, one out of three peptides which is 'QHALSVGPQTTMLSVR' is a specific biomarker while the other two are not. For collagen type 18 alpha 1, both peptides found which are 'QPGLPGPKVR' and 'GLPGPKVR' are specific to the porcine peptide marker. Lastly, for collagen type 15 alpha 1, only one peptide was determined which is 'LPGPPGPPGQPGLPGSR', yet it is not a specific peptide marker for porcine. All non-specific peptide biomarkers were determined by which organism they belong to besides porcine as stated in Supplementary Table 3.

This study reveals that *in silico* protein sequence approaches allow for the rapid screening of proteotypic peptides that can be utilised as species biomarkers efficiently. Nonetheless, these peptides require further verification and validation. Indeed, specific peptides are a promising substitute for existing procedures for species authentication and adulteration. Nevertheless, future research can be conducted through analysis like tandem mass spectrometry etc., to strengthen the existence of these mentioned proteotypic peptides that can be useful in the halal authentication field in preventing issues like adulteration and unethical false labelling thus, thus benefit all parties including industry player and consumers, particularly Muslims, Judaism and Hinduism.

CONCLUSION

To deduce, the identification of peptide biomarkers has led to unique peptide sequences that are in agreement with the predetermined characteristics. These can be utilized for the origin organism identification of gelatin and collagen that can distinguish from other species. The use of open-source tools like Trans-Proteomic Pipeline (TPP) that provide a single environment, and comprehensive mechanisms enable to facilitate the analysis of MS/MS data representation, visualisation and peptide identification. Together with sequence alignment, both approaches were able to identify the specific peptide sequences of porcine where by using the same approaches described, other species would be attainable. To recapitulate this study, a total of 69 fasta files have been compiled composed of four different species, multiple sequence alignments of 10 collagen proteins have been conducted and 197 peptide sequences were discovered of which 81 the sequences belong to collagen and the rest 118 own by gelatin.

ACKNOWLEDGEMENTS

The author would like to acknowledge MITRANS Vanguard Fund and Journal Support Fund, IPSIS UiTM for funding the study and publication fees respectively.

ETHICAL STATEMENT

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Amir, S.H., Yuswan, M.H., Aizat, W.M., Mansor, M.K., Desa, M.N.M., Yusof, Y.A., Song, L.K. & Mustafa, S. 2021. Comparative database search engine analysis on massive tandem mass spectra of pork-based food products for halal proteomics. Journal of Proteomics, 241: 2-7. https://doi.org/10.1016/j.jprot.2021.104240

Ballin, N.Z., Vogensen, F.K. & Karlsson, A.H. 2009. Species determination - Can we detect and quantify meat adulteration? Meat Science, 83(2): 165-174. https://doi.org/10.1016/j.meatsci.2009.06.003

Buckley, M. 2016. Species identification of bovine, ovine and porcine type 1 collagen; Comparing peptide mass fingerprinting and LC-based proteomics methods. International Journal of Molecular Sciences, 17(4): 445. https://doi.org/10.3390/ijms17040445

Calvano, C.D., Monopoli, A., Loizzo, P., Faccia, M. & Zambonin, C. 2013. Proteomic approach based on MALDI-TOF MS to detect powdered milk in fresh cow's milk. Journal of Agricultural and Food Chemistry, 61(8): 1609-1617. https://doi.org/10.1021/jf302999s

Cebi, N., Dogan, C.E., Mese, A.E., Ozdemir, D., Arici, M. & Sagdic, O. 2019. A rapid ATR-FTIR spectroscopic method for classification of gelatin gummy candies in relation to the gelatin source. Food Chemistry, 277: 373-381. https://doi.org/10.1016/j.foodchem.2018.10.125

Cifuentes, A. 2017. Foodomics, foodome and modern food analysis. TrAC Trends in Analytical Chemistry, 96: 1. https://doi.org/10.1016/j.trac.2017.09.001

Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J.K., Martin, D.B., Nesvizhskii, A.I. & Aebersold, R. 2010. A guided tour of the Trans-Proteomic Pipeline. Proteomics, 10(6): 1150-1159. https://doi.org/10.1002/pmic.200900375

Gallardo, J.M., Ortea, I. & Carrera, M. 2013. Proteomics and its applications for food authentication and food-technology research. TrAC Trends in Analytical Chemistry, 52: 135-141. https://doi.org/10.1016/j.trac.2013.05.019

Grundy, H.H., Reece, P., Buckley, M., Solazzo, C.M., Dowle, A.A., Ashford, D., Charlton, A.J., Wadsley, M.K. & Collins, M. J. 2016. A mass spectrometry method for the determination of the species of origin of gelatine in foods and pharmaceutical products. Food Chemistry, 190: 276-284. https://doi.org/10.1016/j.foodchem.2015.05.054

Hashim, D.M., Man, Y.B.C., Norakasha, R., Shuhaimi, M., Salmah, Y. & Syahariza, Z. A. 2010. Potential use of Fourier transform infrared spectroscopy for differentiation of bovine and porcine gelatins. Food Chemistry, 118(3): 856-860. https://doi.org/10.1016/j.foodchem.2009.05.049

Husin, M.M. 2022. Gula-gula bergelatin babi tular semula. Sinar Harian [WWW Document]. Sinar Harian. URL https://www.sinarharian.com.my/article/215261/berita/semasa/gula-gula-bergelatin-babi-tular-semula (accessed 12.19.2023).

Jagadeesh, D.S., Kannegundla, U. & Reddy, R.K. 2017. Application of proteomic tools in food quality and safety. Advances in Animal and Veterinary Sciences, 5(5): 213-225.

Jannat, B., Ghorbani, K., Kouchaki, S., Sadeghi, N., Eslamifarsani, E., Rabbani, F. & Beyramysoltan, S. 2020. Distinguishing tissue origin of bovine gelatin in processed products using LC/MS technique in combination with chemometrics tools. Food Chemistry, 319: 2-6. https://doi.org/10.1016/j.foodchem.2020.126302

Jaswir, I., Mirghani, M.E.S., Salleh, H.M., Ramli, N., Octavianti, F. & Hendri, R. 2016. An overview of the current analytical methods for halal testing. In: Contemporary Issues and Development in the Global Halal Industry. Springer, Singapore. pp. 291-300. https://doi.org/10.1007/978-981-10-1452-9_27

Kariduraganavar, M.Y., Kittur, A.A. & Kamble, R.R. 2014. Polymer synthesis and processing. In: Natural and Synthetic Biomedical Polymers. G. Sangamesh, Kumbar, T. Cato, Laurencin and M. Deng (Eds.). Elsevier, Oxford. pp. 1-31. https://doi.org/10.1016/B978-0-12-396983-5.00001-6

Korte R. & Brockmeyer, J. 2017. Novel mass spectrometry approaches in food proteomics. TrAC Trends in Analytical, 96: 99-106. https://doi.org/10.1016/j.trac.2017.07.010

Manning, L. & Soon, J.M. 2016. Food safety, food fraud, and food defense: A fast evolving literature. Journal of Food Science, 81(4): 823-834. https://doi.org/10.1111/1750-3841.13256

Nugraha, R., Rahayu, F. & Nurilmala, M. 2022. Identification of peptide biomarkers for halal gelatin using bioinformatics techniques. IOP Conference Series: Earth and Environmental Science, 10333: 2-7. https://doi.org/10.1088/1755-1315/1033/1/012063

Orduna, R., Husby, E., Yang, C.T., Ghosh, D. & Beaudry, F. 2015. Assessment of meat authenticity using bioinformatics, targeted peptide biomarkers and high-resolution mass spectrometry. Food Additives & Contaminants: Part A, 32(10): 1709-1717. https://doi.org/10.1080/19440049.2015.1064173

Ortea, I., O'Connor, G. & Maquet, A. 2016. Review on proteomics for food authentication. Journal of Proteomics, 147: 212-225. https://doi.org/10.1016/j.jprot.2016.06.033

Sha, X.M., Hu, Z.Z., Ye, Y.H., Xu, H. & Tu, Z.C. 2019. Effect of extraction temperature on the gelling properties and identification of porcine gelatin. Food Hydrocolloids, 92: 163-172. https://doi.org/10.1016/j.foodhyd.2019.01.059

Sha, X.M., Zhang, L.J., Tu, Z.C., Zhang, L.Z., Hu, Z.Z., Li, Z., Li, X., Huang, T., Wang, H., Zhang, L. & Xiao, H. 2018. The identification of three mammalian gelatins by liquid chromatography-high resolution mass spectrometry. Lwt Food science and Technology, 89: 74-86. https://doi.org/10.1016/j.lwt.2017.10.001

Silva, T.H., Moreira-Silva, J., Marques, A.L., Domingues, A., Bayon, Y. & Reis, R.L. 2014. Marine origin collagens and its potential applications. Marine Drugs, 12(12): 5881-5901. https://doi.org/10.3390/md12125881

Silvipriya, K., Kumar, K., Bhat, A., Kumar, B., John, A. & Lakshmanan, P. 2015. Collagen: Animal

Sources and Biomedical Application. Journal of Applied Pharmaceutical Science, 5(3): 123-127. https://doi.org/10.7324/JAPS.2015.50322

Stadler, R.H., Tran, L.A., Cavin, C., Zbinden, P. & Konings, E.J. 2016. Analytical Approaches to Verify Food Integrity: Needs and Challenges. Journal of AOAC International, 99(5): 1135-1144. https://doi.org/10.5740/jaoacint.16-0231

Tighe, P.J., Ryder, R.R., Todd, I. & Fairclough, L.C. 2015. ELISA in the multiplex era: Potentials and pitfalls. Proteomics Clinical Application, 9(3-4): 406-422. https://doi.org/10.1002/prca.201400130

Yang, C.T., Ghosh, D. & Beaudry, F. 2018. Detection of gelatin adulteration using bio-informatics, proteomics and high-resolution mass spectrometry. Food Additives and Contaminants: Part A, 35(4): 599-608. https://doi.org/10.1080/19440049.2017.1416680

Yilmaz, M.T., Kesmen, Z., Baykal, B., Sagdic, O., Kulen, O., Kacar, O., Yetim, H. & Baykal, A.T. 2013. A novel method to differentiate bovine and porcine gelatins in food products: nanoUPLC-ESI-Q-TOF-MS(E) based data independent acquisition technique to detect marker peptides in gelatin. Food Chemistry, 141(3): 2450-2458. https://doi.org/10.1016/j.foodchem.2013.05.096

Yuswan, M.H., Jalil, N.H.A., Mohamad, H., Keso, S., Mohamad, N.A., Yusoff, T.S.T.M., Ismail, N.F., Manaf, Y.N.A., Hashim, A.M., Desa, M.N.M., Yusof, Y.A. & Mustafa, S. 2021. Hydroxyproline determination for initial detection of halal-critical food ingredients (gelatin and collagen). Food Chemistry, 337: 127762. https://doi.org/10.1016/j.foodchem.2020.127762

Zhang, Q., Tu, Z., Wang, H., Huang, X., Shi, Y., Sha, X. & Xiao, H. 2014. Improved glycation after ultrasonic pretreatment revealed by high-performance liquid chromatography-linear ion trap/Orbitrap high-resolution mass spectrometry. Journal of Agricultural and Food Chemistry, 62(12): 2522-2530. https://doi.org/10.1021/jf5002765

Zhang, T., Shi, X. & Netto, J. 2019. Determination of Animal Species Origin from Gelatin in Food and Pharmaceutical Products by LC-MS/MS [WWW Document]. Sciex. URL https://sciex.com/content/dam/SCIEX/pdf/posters/Food-Fraud-Authenticity.pdf (accessed 12.19.2023)