HALALSPHERE
International Islamic University Malaysia - INHART

Incorporation of Partial Least Squares-Discriminant Analysis with Ultra-High-Performance Liquid Chromatography Diode-Array Detector for Authentication of Skin Gelatine Sources

Muhamad Shirwan Abdullah Sani[a,b*], Azilawati Mohd Ismail[c], Azman Azid[d], Mohd Saiful Samsudin[e] and Mohd Hafis Yuswan[f].

[a]International Institute for Halal Research and Training (INHART), Level 3, KICT Building, International Islamic University Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
[b]The Catalytixs Solutions, No. 713, Jalan DPP ¼, Desa Permai Pedas, 71400 Pedas, Negeri Sembilan, Malaysia.
[c]Malaysia Halal Analysis Centre (MyHAC), No.1, Persiaran Teknologi 1, Lebuh Enstek, 71760 Bandar Baru Enstek, Negeri Sembilan, Malaysia.
[d]Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin (UniSZA), 22200 Besut, Terengganu, Malaysia.
[e]Environmental Technology Division, School of Industrial Technology, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
[f]Halal Products Research Institute, Putra Infoport, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
*Corresponding author: shirwansany@iium.edu.my

Received: 10/8/2023
Accepted: 14/12/2023
Published: 31/1/2024

Keywords:
PCA; PLS-DA; UHPLC-DAD; Amino acids; Skin gelatine

Abstract

This research seeks to (1) authenticate sources of skin gelatine by combining putative 17 amino acids (AAs) analysis with chemometrics by Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD) and (2) create AA profiles in skin gelatines. The classification capability of partial least square-discriminant analysis (PLS-DA) models was assessed to determine the most effective discriminant model. Principal component analysis (PCA) with quartimax rotation was utilised to accurately organise gelatine clusters and assign the significantly contributing AAs to each cluster. The PLS-DA model with 13 AAs (PLS-DAVIPAA) outperformed the PLS-DA model with 17 AAs (PLS-DAAA) because its R^2Y (0.938), R^2X (0.881), and Q^2 (0.929) values were greater. With 13 significant AAs, the PLS-DAVIPAA model obtained cluster classification accuracy of 100% on training and cross-validation datasets and 93.3% on testing and verification datasets. The chemical structure of gelatines may shed light on the interactions between AAs. Following six quartimax rotations, the gelatines were grouped correctly. The PCA showed the dominant presence of these AAs: L-Valine, L-Phenylalanine and L-Tyrosine in porcine gelatine; Glycine, L-Threonine, L-Arginine, L-Methionine, L-Histidine and L-Serine in fish gelatine; and L-Hydroxyproline, L-Leucine and L-Proline in bovine gelatine. The authority could use this technique to set a standard for authenticating skin gelatine samples.

1. Introduction

The gelatine industry was worth $1.34 billion in 2019. Approximately 80% of hydrolysed skin gelatine produces food gelatine products, such as marshmallows, gummies, and candies (Yap & Gam, 2019). Since most skin gelatines originate from porcine and bovine (Rather et al., 2022), some gelatine manufacturers commit food fraud via fraudulent adulteration, substitution and mislabeling of the source (Zia et al., 2020) due to the high demand and escalated cost to produce gelatine from various sources. This action was evident in previously reported food fraud on marshmallows, gumdrops, and Turkish delights containing declared halal gelatine (Garcia-Vaquero & Mirzapour-Kouhdasht, 2023). It could undermine the customers' right to consume gelatine-based products, especially for populations with certain religious beliefs. Our work addresses this action by establishing a method for authenticating skin gelatine sources, specifically porcine, bovine, and fish. This work used an Ultra-High-Performance Liquid Chromatography-Diode-Array Detector (UHPLC-DAD) and chemometrics to identify unique amino acids (AAs) as biomarkers in each type of skin gelatine.

In addition to UHPLC-DAD, Liquid Chromatography-Mass Spectrometer (LC/MS) and Liquid Chromatography Time-of-Flight Mass Spectrometer (LC-QTOF/MS) have been used to validate skin gelatine sources. Despite their high sensitivity, these mass spectrometer-based methods are costly (Rohman et al., 2020). Azilawati et al. (2015) used a High-Performance Liquid Chromatography-Fluorescence Detector (HPLC-FLD) in conjunction with principal component analysis (PCA) to measure amino acids (AAs) instead of LC/MS and LC-QTOF/MS. Yuswan et al. (2021) used HPLC-FLD to separate 17 AAs, validating and verifying the AA method and applying it to gelatine and collagen. Based on these works, we used this less expensive method in our study, except that we detected the separated AAs with a diode array detector, which has been shown to authenticate skin gelatine sources. This method was validated and verified on marshmallows, with the results published by Sani et al. (2021).

Previous studies validated and verified analytical methods for detecting AAs by establishing performance characteristics such as calibration curve, linearity range, the limit of detection and

quantification, sensitivity, and accuracy (Sani et al., 2020). Yuswan et al. (2021) performed a similar approach to comply with the requirement of ISO 17025: 2017 (Yuswan et al., 2020). However, Idris et al. (2021) proposed using putative biomarkers as an alternative and more expedient method for halal authentication, particularly in the case of ad-hoc analysis, due to the cost and time involved in performing these steps to achieve acceptable performance characteristics. In addition, a negligible number of reports mentioned this technique combined with chemometrics to authenticate skin gelatine sources. Consequently, our study utilised this method for the same objective.

Since it is the most prevalent chemometrics classification technique, discriminant analysis (DA) is utilised extensively in food authentication. Sani et al. (2021) also authenticated gelatine products using DA, whereas there are few reports of PLS-DA applications for authenticating gelatine sources. (Sani et al., 2020) noted that PLS-DA characteristics such as Fisher statistics and p-value could serve as discriminatory indicators for authentication purposes. In contrast, Sharin et al. (2021) stated that the importance of the variable in the projection (VIP) from the PLS-DA could identify the significantly contributing variables to the classification functions. Most previously reported chemometrics for food applications also conducted an authentication study at a significance level (α) of 0.05, allowing for a 5% error. Nevertheless, since the contamination of food with non-halal ingredients is a sensitive issue for Muslim consumers, our study authenticated the gelatine at α of 0.01 to increase the reliability of the results. The absence of employing Fisher statistics, p-value and α of 0.01 for halal authentication may impede the development of an authentication method for skin gelatine sources relative to other authentication methods, as these criteria have been shown to reduce the dimensionality of fatty acids methyl ester, thereby facilitating the authentication of oil-based products (Idris et al., 2021). Due to these considerations, our study explored the classification ability of PLS-DA by examining its features on a dataset of skin gelatine.

The next step in analysing the distribution of AAs in each source of skin gelatine is to use exploratory principal component analysis (PCA). Azilawati et al. (2015) and Ismail et al. (2021) performed the PCA. They assigned the contributing AAs by manually removing the skin gelatine samples from the dataset, which is undesirable because it presents a challenge to the testing laboratory, particularly in obtaining repeatability results due to an unstandardised procedure. To achieve correctly clustered samples, Ayerdi & Graña (2014) suggested using PCA with quartimax rotation, and our study used this strategy with varying values of varimax rotation until the skin gelatines were grouped into distinct clusters. This research could assign the contributing AAs to the various clusters and generate AA profiles for each. Because this approach has not been adopted and explained in previous studies, regulatory authorities are likely to be interested in issuing guidelines for the authentication of skin gelatine.

2. Methodology

2.1 Experimental design

The current study entailed skin gelatine preparation and AA analysis in skin gelatines and marshmallows via Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD). The dataset was subjected to dataset pre-processing, PLS-DA and PCA with quartimax rotation. Figure 1 shows the experimental design of the current study.

2.2 Skin gelatine and marshmallow preparation

Cold-water fish (G7041), bovine (G9382) and porcine (G6144) skin gelatines were purchased from Sigma Aldrich, USA. Porcine, bovine, and fish gelatines were prepared to verify the AAs' presence in actual samples. A volume of 240 mL corn syrup (KHH Double Lion, Malaysia), 240 mL of deionised water and 300 g powdered sugar (MSM Malaysia Holdings Berhad, Malaysia) were mixed with 30 g porcine (Great Lakes Gelatin Co., USA). The mixture was homogenised at 200 rpm for 15 min under heating treatment at 80◦C to ensure the porcine gelatines bloom. Further works by Sani et al. (2021) on the preparation of the marshmallows pallets involving bovine (Halagel (M) Sdn. Bhd., Malaysia) and fish (Wani Erat Food Supplies Sdn Bhd., Malaysia) marshmallows before acid hydrolysis was followed.

Approximately 0.2 g gelatines and marshmallow pellets were acid-hydrolysed with 5 mL of 6 N HCL and incubated for 24 h at 110°C. The hydrolysate solution was mixed with 100 pmol/µL L-Aminobutyric acid (Aaba) in a 100 mL volumetric flask, marked up with distilled water and subjected to filtration with 0.45 µm cellulose acetate membrane to produce mixture A.

2.3 Amino acids analysis of skin gelatines and marshmallow

A volume of 10 µL mixture A was derivatised with 20 µL AccQ.Fluor reagent (Waters, Massachusetts, USA) and 70 µL AccQ.Fluor borate buffer (Waters, Massachusetts, USA) at pH maintained between 8.2 – 10.0 to produce mixture B. Then, mixture B was heated for 10 min at 55°C and spiked with 100 pmol/µL L-Aminobutyric acid (Aaba) solution (Waters, the USA) as an internal standard to produce mixture C. A volume of 1 µL mixture C was injected into UHPLC-DAD of Agilent, USA, and eluted by pre-filtered eluant A (AccQ.Tag™ concentrate (WAT052890)) and B (acetonitrile) for AA separation via a gradient elution set-up by Ismail et al. 2021. The AAs were separated at 36°C and 1 mL/min by a Waters AccQ.The tag column (3.9 mm x 150 mm) was subjected to detection by a diode array detector (DAD) at 260 nm.

The putatively detected AA peaks were confirmed by comparing their retention time with an injected mixture of 17 standard AAs and Aaba solutions, as shown in Figure 1. These AAs were selected based on the initial work by Sani et al. (2021). The performance characteristics of the 17 AAs, including method linearity and accuracy, had been validated and verified in our previous studies (Ismail et al., 2021a; Sani et al., 2021). The AA concentration in each skin gelatine was presented in percentage. The ratio of the peak area of AA over Aaba was computed for each gelatine and subjected to dataset pre-processing.

2.4 Dataset pre-processing

Four dataset types were prepared: training, cross-validation, testing and verification (marshmallow) datasets and these datasets were subjected to pre-processing, i.e. outlier treatment, dataset transformation, and dataset adequacy test at a significant level (α) of 0.01 using XLSTAT-Pro (2019) statistical tools (Addinsoft, Paris, France).

The training dataset of 40 porcine, 40 bovine and 40 fish skin gelatines was prepared for the chemometrics. Two types of cross-validation datasets of the PLS-DA model were also prepared: (1) 12 skin gelatines based on a 10-fold cross-

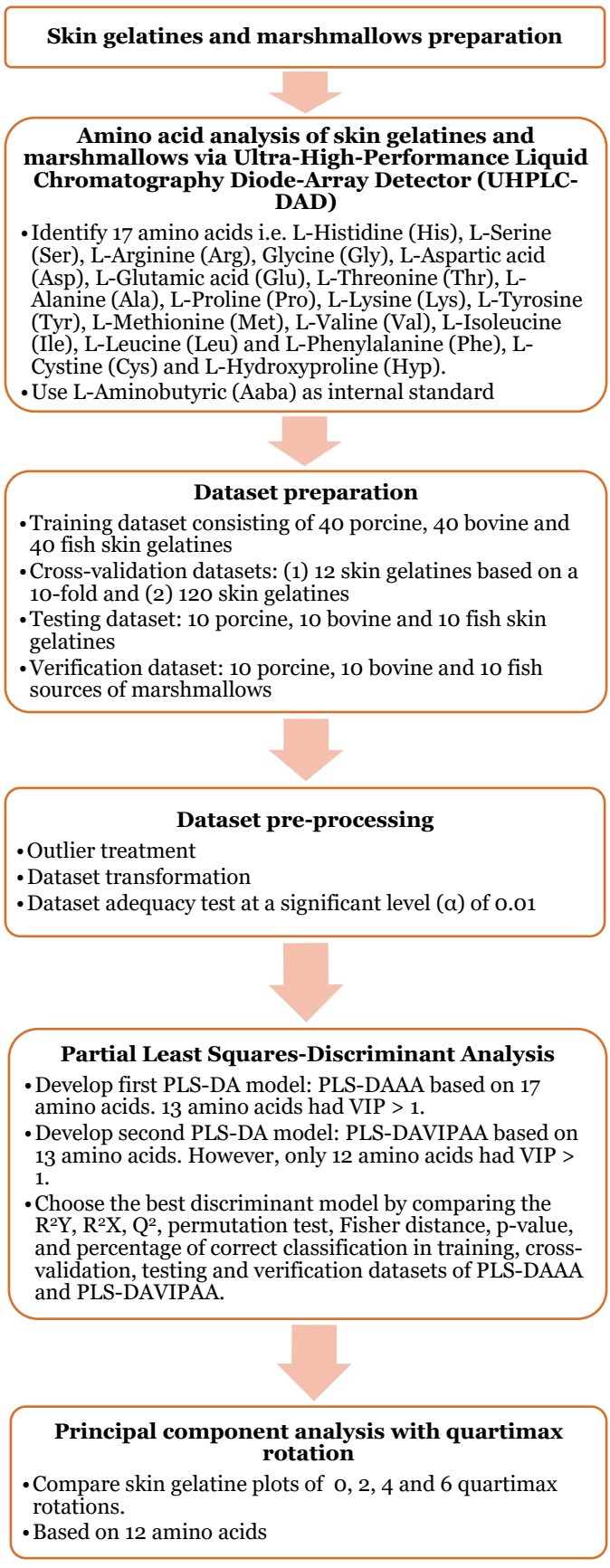

Figure 1: Experimental design of incorporation of partial least squares-discriminant analysis and quartimax rotation with ultra-high-performance liquid chromatography diode-array detector for authentication of skin gelatine sources.

validation dataset (120/10 = 12) (Sharin et al., 2021) and (2) 120 skin gelatines since the PLS-DA feature in XLSTAT-Pro (2019) allows the analysis of cross-validation against the training dataset.

A testing dataset consisting of 10 porcine, 10 bovine and 10 fish skin gelatines was prepared for PLS-DA based on the 80-to-20 minimum ratio of a training-to-testing dataset (Andrada et al., 2015). A verification dataset entailing 10 porcine, 10 bovine and 10 fish sources of marshmallows was also prepared to investigate the accuracy of the PLS-DA models in the actual sample.

The Grubbs and Dixons tests were carried out to identify outliers, replacing them with the mean value of individual AA. The dataset was transformed using the standardise (n-1) method to bring the dataset near normal distribution (Sani et al., 2021).

The dataset adequacy test was carried out using the Kaiser-Meyer-Olkin (KMO) test. The average KMO value was evaluated according to the criteria in the work of Ismail et al. (2021b). The KMO > 0.5 was acceptable for chemometrics.

Post dataset pre-processing, the dataset was subjected to PLS-DA and quartimax rotation of PCA.

2.5 Partial least square-discriminant analysis

The PLS-DA established a discriminating model (DM) for the skin gelatine sources. A new column labelled 'cluster' was added to the training, cross-validation, testing and verification datasets and each gelatine was assigned as 'porcine', 'bovine' and 'fish' clusters. The PLS-DA was carried out twice; the first PLS-DA was carried out involving 17 AAs (PLS-DAAA), while the second PLS-DA (PLS-DAVIPAA) was carried out involving significant AAs with variable importance in the projection () score > 1 (Idris et al., 2022). The significant AAs for PLS-DAVIPAA were selected from the PLS-DAAA's result.

The PLS-DA was carried out at α of 0.01 on the training dataset to establish a DM for the skin gelatines via Equation 1:

$$F(y_i, a_k) = b_0 + \sum_{j=1}^{p} b_i x_{ij} \text{ (Equation 1)}$$

Where F is the function, k is the number of skin gelatine classes for y dependant variable, a is the skin gelatine cluster, b_0 is the DM's intercept, p is the number of AAs, b_i Is DM's coefficient, and x is the observation. The DM discriminated skin gelatines into a class via Equation 2:

$$x = argmax_k F(y_i, a_k) \text{ (Equation 2)}$$

The DM returns the argument (arg) of the F to the maximum (max) class, i.e., k = 3 in this study. The goodness of fit for the DM was evaluated via the R²Y cumulated index, while the quality of the AAs as contributors to the DM was evaluated via the R²X cumulated index. The ability of the DM to classify the skin gelatine clusters was measured via cross-validation of the PLS-DA and represented as a Q² cumulated index.

The classification robustness and accuracy of the PLS-DA models were evaluated using permutation tests with 100 random permutations ($p < 0.01$) via MetaboAnalyst 5.0 (McGill University, Canada). The permutation test calculated the p-value against the hypothesis null (H_0), i.e., no difference among the skin gelatine clusters. The PLS-DA model rejected the H_0

when the p-value was less than 0.01, indicating 99% accuracy in classifying the skin gelatine clusters.

The predictive ability of the DMs was further evaluated on the percentage of correct classification of the porcine, bovine and fish skin gelatines. Cluster dissimilarity was also assessed via the Fisher distance and their distance p-values (Sani et al., 2023). To validate the predictive ability of the DM, the established DM was validated and tested on cross-validation and testing datasets, respectively, and their percentage of correct classifications was evaluated. Based on the R^2Y, R^2X, Q^2, permutation test, Fisher distance, p-value, and percentage of correct classification in training, cross-validation, testing and verification datasets, the classification ability of PLS-DAAA and PLS-DAVIPAA was compared, and the best DM was selected.

2.6 Principal component analysis with quartimax rotation

A Pearson correlation of principal component analysis (PCA) at α of 0.01 was employed to group skin gelatines into porcine, bovine and fish clusters and identify the AA distribution. Twelve significant AAs ($p < 0.01$) identified by PLS-DAVIPAA were transformed and underwent PCA to generate 12 principal components (PCs) known as independent variables. Component score C for y PC number and n sample number as expressed in Equation 3:

$$C_{yn} = f_{y1}a_{n1} + f_{y2}a_{n2} + \cdots + f_{yi}a_{ni} \quad \text{(Equation 3)}$$

Where f is the factor loading (FL), a is the AA concentration, and i is the total number of AA.

Cumulative variability (CV) of two-dimensional PCs entailing PC1 and PC2 were computed for the AA profile exploratory. To group the skin gelatines into porcine, bovine and fish clusters, quartimax rotation at zero (no rotation), two, four and six rotations were carried out. The quartimax rotation was stopped at six rotations as the skin gelatine sources had been clustered with the highest CV. The comparison among the numbers of quartimax rotation was depicted as skin gelatine plots.

The strong, moderate and weak contributing AAs to the clusters were evaluated based on factor loading (FL) criteria: $FL \geq |0.750|$ for strong, $|0.500| < FL < |0.749|$ for moderate and $FL \leq |0.499|$ for weak contributing AAs. The AA profile in each cluster was assessed via a biplot of skin gelatines and amino acids with six quartimax rotations.

3. Results and discussion

3.1 Amino acids content in porcine, bovine and fish skin gelatines

This study investigated the distribution of AA content in porcine, bovine and fish skin gelatines. Table 1 shows the amino acid content in each skin gelatine. The presence of 17 AAs in the gelatine was confirmed with a retention time of SS. The chromatograms and validation and verification of the 17 AAs' performance characteristics, including method linearity and accuracy, were retrievable in our previous studies (Ismail et al., 2021a; Sani et al., 2021). Glycine was dominant, while His was undetected in the porcine skin gelatine. The ranking of AA concentration in the porcine skin gelatine was as follows: Gly (33.66%) > Pro (12.16%) > Hyp (10.63%) > Ala (9.77%) > Glu (6.54%) > Arg (6.30%) > Lys (3.92%) > Asp (3.48%) > Ser (3.08%) > Leu (2.46%) > Val (2.37%) > Thr (1.79%) > Phe (1.56%) > Ile (1.05%) > Met (0.77%) > Tyr (0.45%) > His (0.00%). In comparison with Hafidz & Yaakob (2011), Azilawati et al. (2015) and Widyaninggar et al. (2012), our study had a similar AA distribution: Gly > Pro, Asp > Ser, and Ile > Met > Tyr. The AA distribution in porcine skin gelatine analysed by a validated and verified method by Sani et al. (2021) showed a similar distribution. Although the porcine bone could also be used to produce gelatine, to the authors' knowledge, no report was found on the AA distribution from the porcine bone gelatine.

Based on the ranking of bovine skin gelatine, i.e. Gly (33.83%) > Pro (11.90%) > Hyp (10.89%) > Ala (9.95%) > Glu (6.72%) > Arg (5.95%) > Lys (3.84%) > Asp (3.65%) > Ser (3.11%) > Leu (2.46%) > Val (2.23%) > Thr (1.79%) > Phe (1.47%) > Ile (1.26%) > Met (0.65%) > Tyr (0.28%) > His (0.00%) (Table 1), the AA distribution was similar to the AA distribution of porcine skin gelatine probably due to both of porcine and bovine are mammals. This similarity may render difficulty in differentiating the porcine and bovine skin gelatines. This AA distribution contradicted the findings of Azilawati et al. (2015) and Hafidz & Yaakob (2011) except Gly > Pro, Asp > Ser and Ile > Met > Tyr > His distributions. Valipour et al. (2008) identified the AA distribution of bovine bone gelatine as follows: Gly (17.24%) > Glu (15.56%) > Asp (11.47%) > Pro (9.4%) > Ala (6.67%) > Lys (3.78%) > Thr (3.15%) > Phe (3.15%) > Ser (2.94%) > Arg (2.38%) > Leu (2.27%) > Val (2.09%) > Ile (1.15%) > Met (0.78%) > His (0.67%) > Tyr (0.66%). From this distribution, both bovine skin and bone gelatines had Gly as the dominant AA, similar to the Ile > Met distribution. Conversely, Valipour et al. (2008) identified 0.67% His in the bovine bone gelatine, while our result found no His in the bovine skin gelatine.

Table 1 also presents the AA distribution of fish skin gelatine, which followed this ranking: Gly (35.44%) > Ala (9.73%) > Pro (9.43%) > Arg (6.77%) > Glu (6.25%) > Hyp (6.22%) > Ser (6.02%) > Lys (3.81%) > Asp (3.79%) > Thr (2.76%) > Val (2.02%) > Leu (2.02%) > Met (1.81%) > Phe (1.43%) > Ile (1.20%) > His (0.96%) > Tyr (0.33%). The AA distribution of yellowfin tuna (*Thunnus albacares*) skin gelatine was in line with our finding at the Pro > Ar > Glu and Ile > His > Tyr ranking (Nurilmala et al., 2019). Nevertheless, Nawaz et al. (2020) stated that cold-water fish skin gelatine had lower Hyp than the skin gelatine of warm-water fish. This finding was supported by a higher Hyp in tilapia (*Oreochromis mossambicus*), yellowfin tuna (*Thunnus albacares*) and blackcarp (*Mylopharyngodon piceus*) than cod (*Gadus morhua*), hake (*Merluccius capensis*) and alaska pollock (*Gadus chalcogrammus*). Of the 17 AAs, our fish skin gelatines had similar AA distribution of Met > Phe > Ile > His > Hyl > Tyr in cod; Gly > Ala > Pro, Glu > Hyp, Thr > Val, and Met > Phe in hake; and: Gly > Ala > Pro and Met > Phe > Ile > His > Hyl > Tyr (Derkach et al., 2020). The bone gelatine of *Ephinephelus sp*. has Gly > Pro > Glu > Ala > Arg > Asp > Leu > Ser > Lys > Thr >Val > Phe > Ile > His > Tyr distribution where only Thr > Val and Phe > Ile > His distribution were similar to our study (Suprayitno, 2019). These findings indicated that cold-water fish's skin and bone gelatine had their individual AA distribution, although some similarities are recorded. For this reason, the skin gelatine of porcine, bovine, and fish was differentiated via statistical analysis.

The ANOVA test in Table 1 shows a significant difference in the mean value of AA among the skin gelatines of porcine, bovine, and fish, where skin gelatines with different superscript alphabets were significantly different ($p < 0.01$). The Arg, Pro,

Tyr, Met, Val and Ile significantly differed among the three skin gelatines. Specifically, the gelatine of porcine skin had the highest Pro, Tyr and Val content and the lowest Ile content. The gelatine of bovine skin had the highest Ile percentage and the lowest Arg, Tyr and Met. The fish skin gelatine had the highest Arg and Met content and the lowest Pro and Val content.

The Arg, Pro, Tyr, Met, Val and Ile could be utilised to differentiate porcine and bovine skin gelatines, although the AA distribution between these skin gelatines was similar. However, the content differences between porcine and fish skin gelatines were significant in all AAs except Ala and Lys. Likewise, our study observed a significant difference in all AAs except Asp, Ala, Lys, and Phe in bovine and fish skin gelatines. Nevertheless, the application of the ANOVA test was insufficient to discriminate the three gelatines since more than one AA characterised the gelatines; hence, Sani et al. (2021) and Ismail et al. (2021a) proposed the chemometric application to discriminate the skin gelatines.

3.2 Outlier treatment and dataset adequacy

Prior to the chemometrics, the skin gelatine datasets underwent pre-processing to ensure the dataset fulfilled the chemometrics prerequisite, including outlier treatment, dataset transformation, and dataset adequacy test (Ismail et al., 2021b). The training dataset had 29, 12 and 21 outliers in the porcine, bovine and fish skin gelatines (Table 1), respectively, where our method replaced the outliers with the mean value of each AA (Sani et al., 2021). Then, this training dataset was transformed via the standardised (n-1) method. Although only negligible reports carried out dataset transformation in their works, our study performed the transformation to fulfil the prerequisite of chemometrics (Azilawati et al., 2015). Additionally, various dataset transformations are available for chemometrics, e.g., standardise (n), standard deviation^{-1} (n-1), standard deviation^{-1} (n), centre, 0 to 1 rescaling, 0 to 100 rescaling, Pareto and log methods (Ismail et al., 2021b); however, our study adopted standardised (n-1) as proposed by Sani et al. (2021) for gelatine matrix since high AA numbers achieved normality post this transformation.

Table 1 shows the individual KMO value for each AA, where Met (0.9274) and Ile (0.6075) had the highest and lowest KMO values, respectively. A comparison of the average KMO value (0.7874) with the guideline from Williams and Brown's (2012) study indicated that the dataset adequacy fell on the good ranking (0.7 < KMO < 0.8). Yuswan et al. (2021) and Azilawati et al. (2015) employed chemometrics without declaring the fulfilment of the dataset adequacy; hence, comparing the results may not be possible. Nevertheless, other gelatine studies found that KMO > 0.7 signified that the dataset was adequate for chemometrics (Sani et al., 2021). Our KMO value (0.7874) was higher than the gelatine study by Ismail et al. (2021) (KMO = 0.7542). Based on these comparisons, our KMO value indicated that the dataset was adequate for chemometrics.

3.3 Development of a model of partial least square discriminant analysis for skin gelatine sources

In this study, the partial least square-discriminating analysis (PLS-DA) model generated two components to explain the classification ability of the sources of skin gelatine. Table 2 shows two PLS-DA models, i.e., the PLS-DA model for 17 AAs (PLS-DAAA) and the PLS-DA model for AA with variable importance in the projection (VIP) score > 1 (PLS-DAVIPAA). Sharin et al. (2021) recommended a VIP score > 1 since a high VIP score could explain most of the variance among the porcine, bovine and fish gelatines. The best PLS-DA model was selected by evaluating the performance of the PLS-DAAA and PLS-DAVIPAA models.

The quality of both PLS-DA models was evaluated via R^2Y, R^2X and Q^2 cumulated indices on each component. The R^2Y (0.938) and R^2X (0.881) of PLS-DAVIPAA were higher than the PLS-DAAA (R^2Y = 0.894 and R^2X = 0.771), indicating that the PLS-DAVIPAA was better in explaining the gelatine clusters and AA contribution to the gelatine clusters, respectively. These findings were associated with the definition of R^2Y, which is a sum of determination coefficients between the gelatine clusters and two components. The R^2Y also indicated the variance proportion of the gelatine clusters explained by the PLS-DA model, showing the model's goodness of fit. The $R^2Y \sim 1$ represented a perfect goodness of fit. The R^2X was computed as the sum of determination coefficients between the actual and predicted AAs within the two components. The R^2X value indicated the variability of the AAs where the $R^2X \sim 1$ showed the highest quality of the AAs; hence measured the AAs quality as contributors to the PLS-DA model.

The Q^2 of PLS-DAVIPAA (0.929) was also higher than the Q^2 of PLS-DAAA (0.873), signifying that the two components generated by the PLS-DAVIPAA model had a significant contribution to predictive quality for skin gelatine sources. Since Q^2 was a predictive ability evaluation from the cross-validation dataset, it was also an indicator that the predictive ability of PLS-DAVIPAA for the new dataset was better than PLS-DAAA, where $Q^2 \sim 1$ was the perfect prediction.

Table 2 shows that the AAs with VIP scores > 1 significantly contributed to the predictive quality of skin gelatine sources. The PLS-DAAA identified 13 significant AAs with descending VIP scores, i.e., Hyp (1.417) > Met (1.406) > Thr (1.384) > Tyr (1.356) > Ser (1.351) > His (1.343) > Phe (1.278) > Pro (1.228) > Arg (1.207) > Gly (1.182) > Val (1.177) > Leu (1.171) > Ile (1.023) where the Hyp and Ile were the most and least significant AA, respectively. Based on the VIP score, these AAs explain most of the variance among the porcine, bovine and fish skin gelatines. Hence, these AAs could be used to differentiate the skin gelatine sources. Nevertheless, all 13 AAs of the PLS-DAVIPAA except Ile (0.984) yielded VIP scores > 1, confirming the AA significance in discriminating against the gelatine sources (Table 2).

The calculated permutation tests for PLS-DAAA and PLS-DAVIPAA were similar ($p < 0.01$), indicating they could classify the clusters with 99% accuracy, which could be further explained in the correct classification results. Table 2 also exhibits the correct classification of PLS-DAAA on the porcine, bovine and fish gelatines via 17 AAs. The training and 12 gelatines in cross-validation datasets showed 100% total classification of the porcine, bovine and fish skin gelatines (Table 2), indicating the PLS-DAAA was able to discriminate the gelatines at a 99% confidence level using these datasets. The correct classification of the PLS-DAAA model reduced to 98.8%, 93.3% and 93.3% for 120 skin gelatines in cross-validation, testing and verification datasets, respectively, indicating that the classification ability was negatively affected in the actual sample. On the contrary, the PLS-DAVIPAA had improved the correct classification (100%) using the training and 12 and 120 skin gelatines in cross-validation datasets using 13 significant AAs.

Table 1: Amino acid percentage, number of outliers, and sampling adequacy result of porcine, bovine and fish skin gelatines

Amino acid	Retention time, min[1]	Amino acid concentration in testing dataset, %[1,2]			Number of outliers in testing dataset[1,3,4]			Sampling adequacy value[1,5]
		Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
L-Hydroxyproline (Hyp)	1.774	10.63 ± 0.45[a]	10.89 ± 0.67[a]	6.22 ± 0.19[b]	4 (P3, P18, P19 and P40)	1 (B24)	1 (F37)	0.9048
L-Histidine (His)	1.887	0.00 ± 0.01[b]	0.00 ± 0.00[b]	0.96 ± 0.05[a]	3 (P18, P19, P21, P28 and P40)	1 (B29)	0	0.8519
L-Serine (Ser)	2.543	3.08 ± 0.15[b]	3.12 ± 0.15[b]	6.02 ± 0.16[a]	1 (P40)	1 (B24)	1 (F9)	0.7196
L-Arginine (Arg)	2.610	6.30 ± 0.29[b]	5.95 ± 0.25[c]	6.77 ± 0.20[a]	1 (P9)	1 (B24)	1 (F37)	0.6317
Glycine (Gly)	2.776	33.66 ± 1.05[b]	33.83 ± 0.93[b]	35.44 ± 0.88[a]	1 (P9)	1 (B24)	1 (F9 and F39)	0.7586
L-Aspartic acid (Asp)	2.987	3.48 ± 0.36[b]	3.65 ± 0.34[ab]	3.79 ± 0.28[a]	2 (P4, P15)	0	1 (F13 and F30)	0.8428
L-Glutamic acid (Glu)	3.365	6.54 ± 0.48[a]	6.72 ± 0.47[a]	6.25 ± 0.33[b]	1 (P15)	0	1 (F13 and F30)	0.8320
L-Threonine (Thr)	3.720	1.79 ± 0.08[b]	1.79 ± 0.07[b]	2.76 ± 0.08[a]	1 (P9)	1 (B24)	1 (F9)	0.7721
L-Alanine (Ala)	4.077	9.77 ± 0.51[a]	9.95 ± 0.57[a]	9.73 ± 0.33[a]	1 (P15)	0	2 (F13 and F30)	0.7139
L-Proline (Pro)	4.668	12.16 ± 0.32[a]	11.90 ± 0.36[b]	9.43 ± 0.18[c]	2 (P9, P15)	1 (B13)	2 (F13 and F39)	0.8938
L-Lysine (Lys)	6.009	3.92 ± 0.31[a]	3.84 ± 0.40[a]	3.81 ± 0.20[a]	1 (P15)	0	1 (F13 and F30)	0.7901
L-Tyrosine (Tyr)	6.107	0.45 ± 0.03[a]	0.28 ± 0.03[c]	0.33 ± 0.03[b]	2 (P6 and P40)	0	1 (F38)	0.8049
L-Methionine (Met)	6.351	0.77 ± 0.04[b]	0.65 ± 0.18[c]	1.81 ± 0.08[a]	1 (P8)	0	3 (F9, F37 and F38)	0.9274
L-Valine (Val)	6.624	2.37 ± 0.04[a]	2.23 ± 0.06[b]	2.02 ± 0.04[c]	2 (P9 and P15)	1 (B13)	1 (F13)	0.7890
L-Isoleucine (Ile)	7.940	1.05 ± 0.03[c]	1.26 ± 0.03[a]	1.20 ± 0.04[b]	2 (P8 and P15)	1 (B5)	0	0.6075
L-Leucine (Leu)	8.052	2.46 ± 0.04[a]	2.46 ± 0.07[a]	2.02 ± 0.06[b]	2 (P9 and P15)	1 (B13)	1 (F13)	0.7797
L-Phenylalanine (Phe)	8.184	1.56 ± 0.05[a]	1.47 ± 0.10[b]	1.43 ± 0.05[b]	2 (P10 and P11)	2 (B12 and B24)	3 (F8, F26 and F37)	0.6588
Total outliers	nr	nr	nr	nr	29	12	21	nr
Average KMO value	nr	nr	nr	nr	nr	nr	nr	0.7874

Note: [1]nr = not related.

[2]Different superscript alphabets indicated a significant difference in average relative error mean.

[3]Number of detected outliers by Grubbs and Dixon tests.

[4]Skin gelatine in parenthesis indicates the outlier presence.

[5]Sampling adequacy test by Kaiser-Meyer-Olkin (KMO) test.

Table 2: Classification matrix of training, validation and testing datasets of partial least square – discriminant analysis

Discriminating model (DM)	Discriminating model quality		Dataset	Correct classification, %	Number of gelatines, Fisher distance value and p-value of Fisher distance in skin gelatine cluster[3,4]			Total
	R^2Y, R^2X and Q^2 cumulated indices and permutation test ($p < 0.01$)	Ranking of significant amino acid ($p < 0.01$)[1,2]			Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
Partial least square–discriminant analysis for 17 amino acids (PLS-DAAA)	R^2Y : 0.894; R^2X : 0.771; Q^2 : 0.873; and $p < 0.01$	Hyp (1.417) > Met (1.406) > Thr (1.384) > Tyr (1.356) > Ser (1.351) > His (1.343) > Phe (1.278) > Pro (1.228) > Arg (1.207) > Gly (1.182) > Val (1.177) > Leu (1.171) > Ile (1.023)	Training dataset (120 skin gelatines)					
			Porcine gelatine	100	40 (0, 1)	0 (294, < 0.0001)	0 (2764, < 0.0001)	40
			Bovine gelatine	100	0 (294, < 0.0001)	40 (0, 1)	0 (2762, < 0.0001)	40
			Fish gelatine	100	0 (2764, < 0.0001)	0 (2762, < 0.0001)	40 (0, 1)	40
			Total	100	40	40	40	120
			Cross-validation dataset (10-fold – 12 skin gelatines)					
			Porcine gelatine	100	4 (0, 1)	0 (294, < 0.0001)	0 (2764, < 0.0001)	4
			Bovine gelatine	100	0 (294, < 0.0001)	4 (0, 1)	0 (2762, < 0.0001)	4
			Fish gelatine	100	0 (2764, < 0.0001)	0 (2762, < 0.0001)	4 (0, 1)	4
			Total	100	4	4	4	12
			Cross-validation dataset (120 skin gelatines)					
			Porcine gelatine	100	40 (0, 1)	0 (294, < 0.0001)	0 (2764, < 0.0001)	40
			Bovine gelatine	96.5	2 (294, < 0.0001)	38 (0, 1)	0 (2762, < 0.0001)	40
			Fish gelatine	100	0 (2764, < 0.0001)	0 (2762, < 0.0001)	40 (0, 1)	40
			Total	98.8	42	38	40	120
			Testing dataset (30 skin gelatines)					
			Porcine gelatine	90	9 (0, 1)	1 (294, < 0.0001)	0 (2764, < 0.0001)	10
			Bovine gelatine	90	1 (294, < 0.0001)	9 (0, 1)	0 (2762, < 0.0001)	10
			Fish gelatine	100	0 (2764, < 0.0001)	0 (2762, < 0.0001)	10 (0, 1)	10
			Total	93.3	10	10	10	30
			Verification dataset (30 marshmallows)					
			Porcine gelatine	90	9 (0, 1)	1 (294, < 0.0001)	0 (2764, < 0.0001)	10
			Bovine gelatine	90	1 (294, < 0.0001)	9 (0, 1)	0 (2762, < 0.0001)	10
			Fish gelatine	100	0 (2764, < 0.0001)	0 (2762, < 0.0001)	10 (0, 1)	10
			Total	93.3	10	10	10	30
PLS-DA for AA with variable importance in the projection (VIP) > 1 (PLS-DAVIPAA)	R^2Y : 0.938; R^2X : 0.881; Q^2 : 0.929; and $p < 0.01$	His (1.282) > Tyr (1.275) > Met (1.257) > Ser (1.249) > Phe (1.198) > Thr (1.183) > Hyp (1.160) > Pro (1.121) > Arg (1.110) >	Training dataset (120 skin gelatines)					
			Porcine gelatine	100	40 (0, 1)	0 (227, < 0.0001)	0 (2901, < 0.0001)	40
			Bovine gelatine	100	0 (227, < 0.0001)	40 (0, 1)	0 (3316, < 0.0001)	40
			Fish gelatine	100	0 (2901, < 0.0001)	0 (3316, < 0.0001)	40 (0, 1)	40
			Total	100	40	40	40	120
			Cross-validation dataset (10-fold – 12 skin gelatines)					
			Porcine gelatine	100	4 (0, 1)	0 (227, < 0.0001)	0 (2901, < 0.0001)	11
			Bovine gelatine	100	0 (227, < 0.0001)	4 (0, 1)	0 (3316, < 0.0001)	15
			Fish gelatine	100	0 (2901, < 0.0001)	0 (3316, < 0.0001)	4 (0, 1)	14

Discriminating model (DM)	Discriminating model quality		Dataset	Correct classification, %	Number of gelatines, Fisher distance value and p-value of Fisher distance in skin gelatine cluster[3,4]			Total
	R^2Y, R^2X and Q^2 cumulated indices and permutation test ($p < 0.01$)	Ranking of significant amino acid ($p < 0.01$)[1,2]			Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
		Gly (1.102) > Val (1.041) > Leu (1.015) > Ile (0.984)	Total	100	4	4	4	40
			Cross-validation dataset (120 skin gelatines)					
			Porcine gelatine	100	40 (0, 1)	0 (227, < 0.0001)	0 (2901, < 0.0001)	40
			Bovine gelatine	100	0 (227, < 0.0001)	40 (0, 1)	0 (3316, < 0.0001)	40
			Fish gelatine	100	0 (2901, < 0.0001)	0 (3316, < 0.0001)	40 (0, 1)	40
			Total	100	40	40	40	120
			Testing dataset (30 skin gelatines)					
			Porcine gelatine	90	9 (0, 1)	1 (227, < 0.0001)	0 (2901, < 0.0001)	10
			Bovine gelatine	90	1 (227, < 0.0001)	9 (0, 1)	0 (3316, < 0.0001)	10
			Fish gelatine	100	0 (2901, < 0.0001)	0 (3316, < 0.0001)	10 (0, 1)	10
			Total	93.3	10	10	10	30
			Verification dataset (30 marshmallows)					
			Porcine gelatine	90	9 (0, 1)	1 (227, < 0.0001)	0 (2901, < 0.0001)	10
			Bovine gelatine	90	1 (227, < 0.0001)	9 (0, 1)	0 (3316, < 0.0001)	10
			Fish gelatine	100	0 (2901, < 0.0001)	0 (3316, < 0.0001)	10 (0, 1)	10
			Total	93.3	10	10	10	30

Note: [1]Value in parenthesis was an F-statistic value of significant amino acid with variable importance in the projection (VIP) > 1.

[2]Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine, Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine and Phe = L-Phenylalanine.

[3]Values in parenthesis were the Fisher distance value and p-value of Fisher distance, respectively.

[4]Calculated p-value of Fisher distance < 0.01 indicated that the three clusters were significantly different.

The correct classification was also reduced to 93.3% for testing and verification datasets, where the verification dataset was the actual sample (marshmallow). The PLS-DAVIPAA also had a higher and significant Fisher distance value ($p < 0.0001$), i.e., porcine vs bovine (227), porcine vs fish (2901) and bovine vs fish (3316) compared to PLS-DAAA. These results showed that the clusters produced by PLS-DAVIPAA were significantly separated, thus allowing higher authentication accuracy for the skin gelatine sources than the PLS-DAAA. Based on these results, PLS-DAVIPAA with 13 significant AAs was the best DM model to authenticate the skin gelatine sources.

Figure 2 (b) also supported the finding for higher accuracy of PLS-DAVIPAA than PLS-DAAA in Figure 2 (a). The PLS-DAVIPAA showed that each porcine, bovine and fish skin gelatine was located nearer within their clusters than the skin gelatine plot by PLS-DAAA. This result was evident in a narrower range of component 2 score of PLS-DAVIPAA than PLS-DAAA, i.e., porcine skin gelatine (-3 to 0), bovine skin gelatines (1 to 4) and fish skin gelatine (-1 to 1) in Figure 2 (b). In contrast, the component 2 score of the gelatine sources for PLS-DAAA was as follows: -4 to 0 for porcine skin gelatines, 0 to 5 for bovine skin gelatine, and -3 to 1 for fish skin gelatine (Figure 2 (a)). Nevertheless, comparing the clusters via these plots was prone to inaccurate interpretation due to macroscopic evaluation; hence, comparison via R^2Y, R^2X, Q^2, Fisher distance, p-value, percentage of correct classification in training, cross-validation, testing and verification datasets, and permutation test value was preferable. Westerhuis et al. (2008) also recommended against using skin gelatine plots to interpret the classification accuracy.

Of these AAs, Figure 2 (c) depicted the 17 AAs plot from PLS-DAAA with individual values of the correlation matrix (CMV) for each AA, where His, Hyp, Ser, Thr, Met, Pro and Leu had CMV of 0.95 – 0.88 and were followed by Val, Arg and Gly with CMV of 0.69 – 0.63 in component 1. The Ile, Phe, Tyr, Glu, Asp, Lys and Ala had the lowest CMV (0.47 – 0.064). For component 2, the Tyr and Phe had the highest CMV (0.85 – 0.83); Arg, Gly, Val, Lys and Ala had the moderate CMV (0.70 – 0.51); and Glu, Asp, Pro, Leu, Ile, Met, Thr, Ser, His and Hyp had the lowest CMV (0.38 – 0.04). Figure 2 (d) of PLS-DAVIPAA shows the CMV of 13 AAs where the CMV for each AA in components 1 and 2 had a similar value to the PLS-DAAA. Jannat et al. (2018, 2020b) carried out PLS-DA analyses to distinguish porcine, bovine and fish gelatines, but none explained the CMV of the detected compounds or AAs. Nevertheless, Ismail et al. (2021) classified the AAs into strong, moderate and weak factor loading for AAs according to the CMV of AAs from principal component analysis (PCA), not PLS-DA. The CMV was used to delineate the AA relationship among them and assign the AA to the gelatine sources (Ismail et al., 2021b).

Figure 2 (d) exhibits positive correlations based on AA direction proximities: His, Ser, Met and Thr; Gly and Arg; Tyr and Phe; and Leu and Pro. On the contrary, negative correlations of AAs were observed based on their opposite direction: His, Ser, Met and Thr against Hyp; Tyr and Phe against Ile; Val against Ile; and Leu and Pro against Ile. Arg and Gly did not correlate with Ile since their directions were at 90°.

Ismail et al. (2021) proposed that the AA's correlations were due to its polarity side chain; however, our study found that only Met has a non-polar side chain, although it had a positive correlation with His, Ser and Thr. Further generic grouping of AAs, e.g. basic, carboxylic, hydroxylic and hydrophobic, based on the chemical characteristics of Derkach et al. (2020), could not support the AA correlations. The opposite side chains of Gly and Arg and Tyr and Phe also signified that the correlations of AAs were independent of their polarity side chain and generic chemical characteristics. Nevertheless, the backbone of the chemical structure may suggest the reason for the positive correlations among the AAs and vice versa. For instance, Met, His, Ser and Thr, and Gly and Arg shared HO-CO-CNH_2-backbone; Tyr and Phe shared HO-CO-CNH_2-CH_2-benzene backbone while Leu and Pro shared HO-CO- backbone. Furthermore, Hyp and Pro, and Leu and Val were also positioned at close proximity that shared HO-CO-pyrrole and HO-CO-NH_2 backbones, respectively.

To assign the AAs to porcine, bovine and fish skin gelatines, the skin gelatines and AA plots shall be overlaid together where the PLS-DA feature of XLSTAT 2019 could not be provided in this study. However, PCA is preferable since AA and skin gelatine plots are available in the PCA feature that serves as exploratory chemometrics. Hence, this current study carried out the AAs assignment via PCA in the next section.

3.4 Exploring amino acid profile in skin gelatines

The PCA application aimed to (1) explain the distribution of 12 AAs by the PLS-DAVIPAA in porcine, bovine and fish skin gelatines and (2) determine the dominant, moderate and low AA content in each cluster and (3) verify the significant 12 AAs via comparing to the factor loading (FL) of each AA. Only 12 AAs were subjected to the PCA since these were AAs with VIP > 1 that the PLS-DAVIPAA determined. The skin gelatine plots in Figure 3 (a) – (d) had two principal components (PCs) with cumulative variability (CV) of 91.68% for Figure 3 (a) and Figure 3 (b), 88.75% for Figure 3 (c) and 88.88% for Figure 3 (d) that explained the 12 AAs distribution. However, our study could not determine the distribution of the 12 AAs in porcine, bovine, and fish skin gelatines since the clusters were mixed. Figure 3 (a) shows the overlapping of porcine and bovine skin gelatine clusters. Our study carried out the varimax rotation as suggested by Otto (2017), but the porcine and bovine clusters were still overlapping. Ayerdi & Graña (2014) proposed quartimax rotation instead to simplify the PC structure and achieve optimal clusters. Hence, our study applied quartimax rotation at two, four and six rotations to enhance the variance of factor loadings (FLs) of the PC, reducing dimensionality and facilitating the explanation of 12 AAs distribution in each skin gelatine (de Almeida et al., 2020). Of these rotations, the four quartimax rotations in Figure 3 (c) and six quartimax rotations in Figure 3 (d) achieved the correct reposition of all the skin gelatine into their clusters. Nevertheless, the six quartimax rotations had a higher CV at 88.88% than the four quartimax rotations; thus, our study selected the six quartimax rotations as the best one.

Figure 3 (e) assigned the AAs to the three clusters by overlaying the skin gelatine and AA plots to investigate their distribution in each skin gelatine. Figure 3 (e) also depicts the absent information in the PLS-DA, such as the dominant, moderate and low AA content in each cluster. The dominant AAs were as follows: Tyr, Phe and Val in porcine gelatine and Met, Thr, Ser, His, Arg and Gly in fish gelatine since these AAs and the clusters were in the same direction.

This finding aligned with Sani et al. (2021) result. Our study findings were also similar to Azilawati et al. (2015) on Tyr, Met, Thr and Ser distribution in porcine and bovine skin gelatines, respectively. The Pro, Leu and Hyp contents were moderate in porcine and bovine skin gelatines. This moderate content of

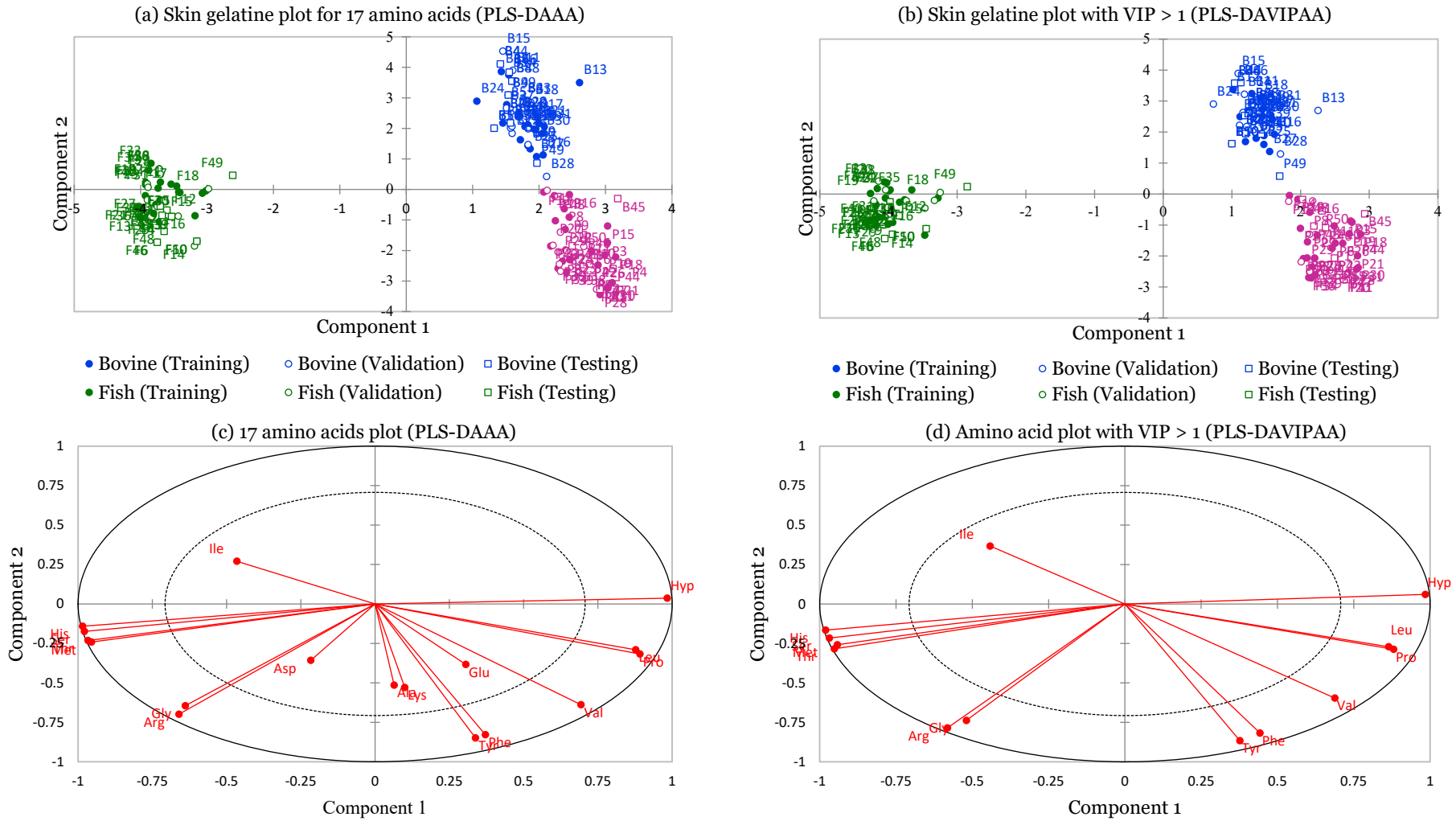

Figure 2: (a) Skin gelatine plot for 17 amino acids (PLS-DAAA), (b) skin gelatine plot with VIP > 1 (PLS-DAVIPAA), (c) 17 amino acids plot (PLS-DAAA) and (d) amino acid plot with VIP > 1 (PLS-DAVIPAA) via partial least square-discriminant analysis.

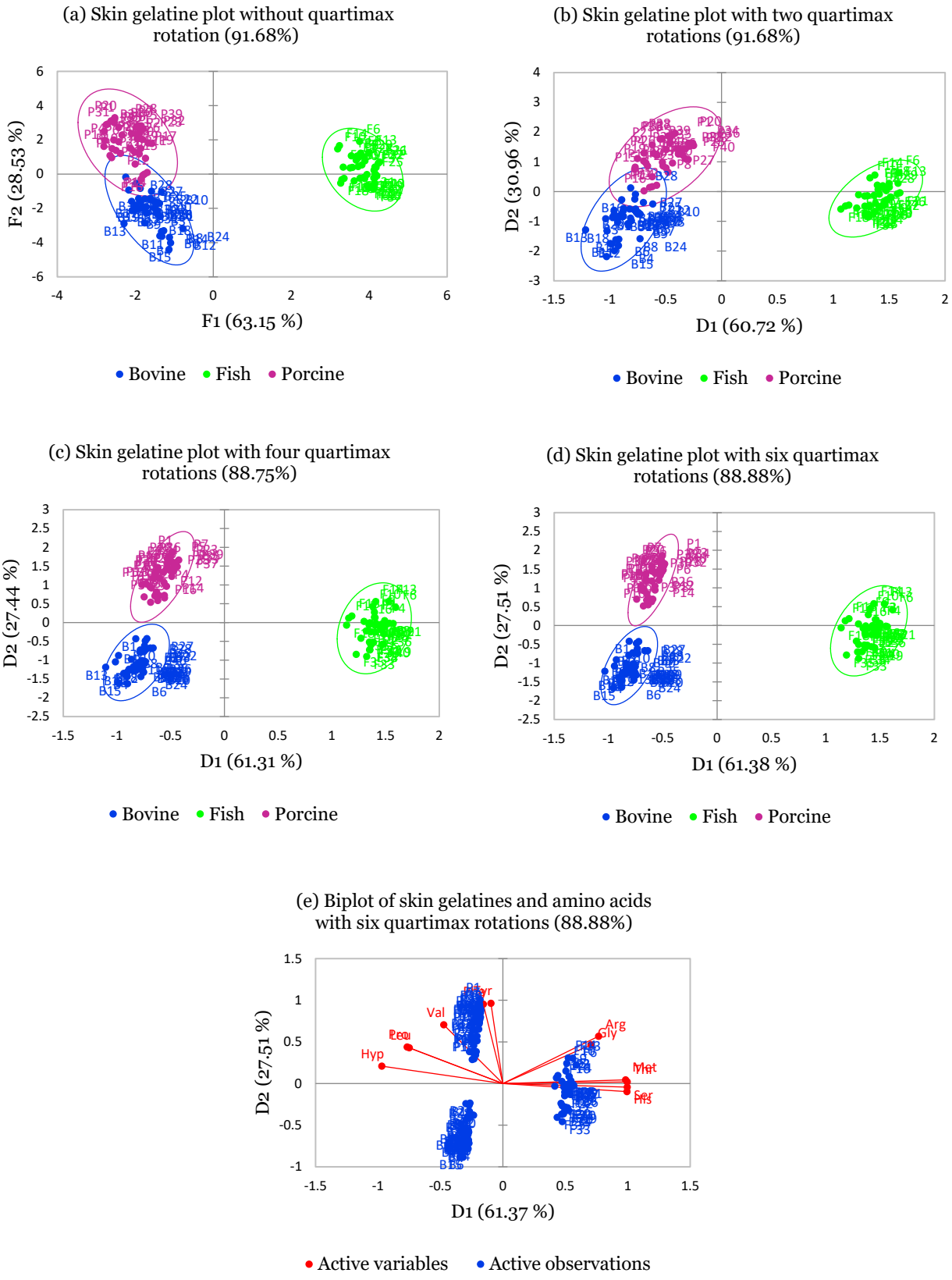

Figure 3: Skin gelatine plots with (a) no quartimax rotation, (b) two quartimax rotations, (c) four quartimax rotations and (d) six quartimax rotations, and (e) biplot of skin gelatines and amino acids with four quartimax rotations.

these AAs was due to their direction between these clusters. Since the Hyp was moderately distributed in porcine and bovine skin gelatines, our result may agree with Yuswan et al. (2021) study that proposed Hyp as one of the biomarkers for halal authentication in gelatine products. The Arg and Gly were low in bovine skin gelatine since their directions were opposite to this cluster. Likewise, fish skin gelatine had low Pro, Leu and Hyp.

Figure 3 (e) also provides information on significant AAs based on the FL of each AA; Hyp, His, Ser, Thr, Pro Tyr, Met, and Leu were the strongly contributing AAs (FL ≥ |0.750|) while Arg, Gly and Val moderately contributing AAs (|0.500| < FL < |0.749|) to the clusters. These results indicated that of the 12 AAs with VIP >1, only 9 were the most significant AAs in this study.

4. Conclusion

This study showed that putatively analysed AAs in skin gelatine via UHPLC-DAD incorporated with DA could discriminate the skin gelatine sources. The DA with a higher percentage of correct classification was superior to PLS-DA for distinguishing skin gelatine sources. The PCA with six quartimax rotations could also assign the skin gelatines to their clusters and provide the AA profile in each cluster. Further study on developing the diagnostic ratio for authentication of skin gelatine sources from this profiling is in the pipeline as a continuation of this study. This study focuses on the classification of skin gelatine sources only since gelatine from this source is the majority used in food industry manufacturing. Hence, a further study including other gelatine sources such as bone and variations, e.g., blooms, will also be carried out. Since this study did not undergo method validation and verification, it may expedite the authentication analysis with less cost and time. Based on this study, the authority may adopt and regulate a standard ad-hoc test to authenticate skin gelatine products.

5. Acknowledgement

The authors would like to acknowledge the Ministry of Higher Education Malaysia for granting Konsortium Institut Halal IPT Malaysia (KIHIM) a research grant 63900911-10205 and Malaysia Halal Analysis Centre (MyHAC) of the Department of Islamic Development Malaysia for the funding, laboratory's facilities, and assistance.

6. Conflict of interest statement

We declare no conflict of interest.

7. Research involving human participants and/or animals

We declare no human and/or animals involved in this study.

8. Informed consent and credit author statement

All authors have granted permission in full knowledge for the publication of this study. The credit author statement as follows: Azilawati Mohd Ismail for methodology; Muhamad Shirwan Abdullah Sani for conceptualisation, data curation and writing of original draft; Azman Azid for software and validation; Mohd Saiful Samsudin and Mohd Hafis Yuswan for visualisation and investigation.

References

Andrada, M. F., Vega-Hissi, E. G., Estrada, M. R., & Garro Martinez, J. C. (2015). Application of k-means clustering, linear discriminant analysis and multivariate linear regression for the development of a predictive QSAR model on 5-lipoxygenase inhibitors. *Chemometrics and Intelligent Laboratory Systems*, *143*, 122–129. https://doi.org/10.1016/j.chemolab.2015.03.001

Ayerdi, B., & Graña, M. (2014). Hybrid extreme rotation forest. *Neural Networks*, *52*, 33–42. https://doi.org/10.1016/j.neunet.2014.01.003

Azilawati, M. I., Hashim, D. M., Jamilah, B., & Amin, I. (2015). RP-HPLC method using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate incorporated with normalisation technique in principal component analysis to differentiate the bovine, porcine and fish gelatins. *Food Chemistry*, *172*, 368–376. https://doi.org/10.1016/j.foodchem.2014.09.093

de Almeida, F. A., Leite, R. R., Gomes, G. F., Gomes, J. H. de F., & de Paiva, A. P. (2020). Multivariate data quality assessment based on rotated factor scores and confidence ellipsoids. *Decision Support Systems*, *129*(October), 113173. https://doi.org/10.1016/j.dss.2019.113173

Derkach, S. R., Voron, N. G., Kuchina, Y. A., & Kolotova, D. S. (2020). *Modified Fish Gelatin as an Alternative to Mammalian Gelatin in Modern Food Technologies*. 27–29.

Garcia-Vaquero, M., & Mirzapour-Kouhdasht, A. (2023). A review on proteomic and genomic biomarkers for gelatin source authentication: Challenges and future outlook. *Heliyon*, *9*(6), e16621. https://doi.org/10.1016/j.heliyon.2023.e16621

Hafidz, R., & Yaakob, C. (2011). Chemical and functional properties of bovine and porcine skin gelatin. *International Food Research Journal*, *18*, 813–817.

Idris, M. H. H., Abdullah Sani, M. S., Mohd Hashim, A., Mohd Zaki, N. N., Abdul Manaf, Y. N., Mohd Desa, M. N., Arshad, S., Yuswan, M. H., Hassan, M. S., Yusof, Y. A., Kamaruddin, M. S., & Mustafa, S. (2022). Forensic feed strategy: Incorporation of multivariate and instrumental analyses for authentication of fish feed sources. *Journal of Halal Industry & Services*, *5*(1), a0000293. https://doi.org/10.36877/jhis.a0000293

Idris, M. H. H., Manaf, Y. N., Desa, M. N. M., Hashim, A. M., Sani, M. S. A., Yusof, Y. A., Kamaruddin, M. S., Mustafa, S., Zaki, N. N. M., Hassan, M. S., & Yuswan, M. H. (2021). A conjunction of sn-2 fatty acids and overall fatty acid composition combined with chemometric techniques increases the effectiveness of lard detection in fish feed. *Chemometrics and Intelligent Laboratory Systems*, *213*(March). https://doi.org/10.1016/j.chemolab.2021.104308

Ismail, A. M., Sani, M. S. A., Azid, A., Zaki, N. N. M., Arshad, S., Tukiran, N. A., Abidin, S. A. S. Z., Samsudin, M. S., & Ismail, A. (2021a). Food forensics on gelatine source via ultra-high-performance liquid chromatography diode-array detector and principal component analysis. *SN Applied Sciences*, *3*(1), 1–19. https://doi.org/10.1007/s42452-020-04061-7

Ismail, A. M., Sani, M. S. A., Azid, A., Zaki, N. N. M., Arshad, S., Tukiran, N. A., Abidin, S. A. S. Z., Samsudin, M. S., & Ismail, A. (2021b). Food forensics on gelatine source via

ultra-high-performance liquid chromatography diode-array detector and principal component analysis. *SN Applied Sciences*, *3*(79), 1–19. https://doi.org/10.1007/s42452-020-04061-7

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*(September 2018). https://doi.org/10.1016/j.foodchem.2020.126302

Jannat, B., Ghorbani, K., Shafieyan, H., Kouchaki, S., Behfar, A., Sadeghi, N., Beyramysoltan, S., Rabbani, F., Dashtifard, S., & Sadeghi, M. (2018). Gelatin speciation using real-time PCR and analysis of mass spectrometry-based proteomics datasets. *Food Control*, *87*, 79–87. https://doi.org/10.1016/j.foodcont.2017.12.006

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36.

Nawaz, A., Li, E., Irshad, S., Xiong, Z., Xiong, H., Shahbaz, H. M., & Siddique, F. (2020). Valorisation of fisheries by-products: Challenges and technical concerns to food industry. *Trends in Food Science and Technology*, *99*, 34–43. https://doi.org/10.1016/j.tifs.2020.02.022

Nurilmala, M., Pertiwi, R. M., Nurhayati, T., Fauzi, S., Batubara, I., & Ochiai, Y. (2019). Characterisation of collagen and its hydrolysate from yellowfin tuna Thunnus albacares skin and their potencies as antioxidant and antiglycation agents. *Fisheries Science*, *85*(3), 591–599. https://doi.org/10.1007/s12562-019-01303-5

Otto, M. (2017). Chemometrics: Statistics and computer application in analytical chemistry. In *Wiley-VCH Verlag GmbH & Co*. https://doi.org/10.1007/s00216-017-0503-7

Rather, J. A., Akhter, N., Ashraf, Q. S., Mir, S. A., Makroo, H. A., Majid, D., Barba, F. J., Khaneghah, A. M., & Dar, B. N. (2022). A comprehensive review on gelatin: Understanding impact of the sources, extraction methods, and modifications on potential packaging applications. *Food Packaging and Shelf Life*, *34*(September), 100945. https://doi.org/10.1016/j.fpsl.2022.100945

Rohman, A., Windarsih, A., Erwanto, Y., & Zakaria, Z. (2020). Review on analytical methods for analysis of porcine gelatine in food and pharmaceutical products for halal authentication. *Trends in Food Science and Technology*, *101*, 122–132. https://doi.org/10.1016/j.tifs.2020.05.008

Sani, M. S. A., Ismail, A. M., Azid, A., & Samsudin, M. S. (2021). Establishing forensic food models for authentication and quantification of porcine adulterant in gelatine and marshmallow. *Food Control*, *130*(December 2020), 108350. https://doi.org/10.1016/j.foodcont.2021.108350

Sani, M. S. A., Nordin, N. F. H., & Elgharbawy, A. A. M. (2023). Halal detection technologies: Analytical method approaches, validation and verification, and multivariate data analysis for halal authentication. In N. N. A. Nizar, S. A. S. Z. Abidin, & A. Bujang (Eds.), *Innovation of Food Products in Halal Supply Chain Worldwide* (pp. 253–271). Elsevier Inc.

Sani, M. S. A., Yuswan, M. H., Desa, M. N. M., & Azid, A. (2020). Procedure of method validation and verification, and multivariate data analysis of chemical measurement for halal-testing laboratories. *Scientific Conference WHS 2019*, *December*, 2–14. https://www.researchgate.net/publication/347489581

Sharin, S. N., Sani, M. S. A., Jaafar, M. A., Yuswan, M. H., Kassim, N. K., Manaf, Y. N., Wasoh, H., Zaki, N. N. M., & Hashim, A. M. (2021). Discrimination of Malaysian stingless bee honey from different entomological origins based on physicochemical properties and volatile compound profiles using chemometrics and machine learning. *Food Chemistry*, *346*(January), 128654. https://doi.org/10.1016/j.foodchem.2020.128654

Suprayitno, E. (2019). Amino acid marshmallow profile from grouper bone gelatin. *International Journal of Scientific and Research Publications*, *9*(11), p9512. https://doi.org/10.29322/ijsrp.9.11.2019.p9512

Valipour, P., Talebian, A., & Ebadi, A. G. (2008). Gelatin preparation from cow's bone by heat-pressure method. *Asian Journal of Chemistry*, *20*(2), 1337–1342.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., Duijnhoven, J. P. M., & Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, *4*(1), 81–89. https://doi.org/10.1007/s11306-007-0099-6

Widyaninggar, A., Triwahyudi, Triyana, K., & Rohman, A. (2012). Differentiation between porcine and bovine gelatin in commercial capsule shells based on amino acid profiles and principal component analysis. *Indonesian Journal of Pharmacy*, *23*(2), 104–109.

Williams, B., & Brown, T. (2012). Exploratory factor analysis : A five-step guide for novices. *Journal of Emergengy Primary Health Care*, *8*(3).

Yuswan, M. H., Nurul, N. H., Mohamad, H., Keso, S., Mohamad, N. A., Tengku, T. S., Ismail, N. F., Abdul Manaf, Y. N., Mohd Hashim, A., Mohd Desa, M. N., Yusof, Y. A., & Mustafa, S. (2021). Hydroxyproline determination for initial detection of halal-critical food ingredients (gelatin and collagen). *Food Chemistry*, *337*(November 2019). https://doi.org/10.1016/j.foodchem.2020.127762

Yuswan, M. H., Sani, M. S. A., Manaf, Y. N. A., & Desa, M. N. M. (2020). Basic requirements of laboratory operation for halal analysis. *KnE Social Sciences*, *2020*, 55–65. https://doi.org/10.18502/kss.v4i9.7315

Zia, Q., Alawami, M., Mokhtar, N. F. K., Nhari, R. M. H. R., & Hanish, I. (2020). Current analytical methods for porcine identification in meat and meat products. *Food Chemistry*, *324*(March), 126664. https://doi.org/10.1016/j.foodchem.2020.126664