SPECIAL GUEST EDITOR SECTION

Development and In-House Validation of an LC–MS and LC–MS/MS Assay for the Determination of Food Fraud for Different Fish Species

Philipp Lasch
University of Applied Sciences Bremerhaven, An der Karlstadt 8, 27568 Bremerhaven, Germany
Steffen Uhlig
QuoData GmbH, Prellerstraße 14, 01309 Dresden, Germany
Carsten Uhlig
Akees GmbH, Ansbacher Straße 11, 10787 Berlin, Germany
Christian Wilhelm, Nicola Bergmann, and Stefan Wittke[1]
University of Applied Sciences Bremerhaven, An der Karlstadt 8, 27568 Bremerhaven, Germany

Background: **Fish and fish products are one of the most important food sources of high commercial interest. The global food trade and the associated risks are constantly presenting new challenges to consumer protection and public authorities, which, among other things, demand state-of-the-art analytical methods to ensure food authenticity.** ***Objective:*** **The establishment of MS-based strategies plays a decisive role alongside the (further) development of ELISA- or DNA-oriented methods.** ***Methods:*** **In the present work, therefore, the development and in-house validation of an LC–MS and LC–MS/MS-based assay for authenticity testing of certain fish species is described.** ***Results:*** **Based on the execution of a validated bottom-up LC–electrospray–MS and MS/MS assay and multivariate analysis, the commercially available species *Lutjanus malabaricus* (red snapper) and *Sebastes* spp. (redfish) are distinguished from each other, whereas an additional 68 samples [nine additional marine species such as pangasius (*Pangasianodon hypophthalmus*), salmon (*Salmo salar*), turbot (*Scophthalmus maximus*), plaice (*Pleuronectes platessa*), sole (*Solea solea*), lemon sole (*Glyptocephalus cynoglossus*), halibut (*Reinhardtius hypoglossoides*), red salmon (*Oncorhynchus nerka*), and great scallop (*Pecten jacobaeus*)] served as blinded negative controls to ensure the specificity of the assay.** ***Conclusions and Highlights:*** **A promising LC–MS and LC–MSMS based assay has been developed that could enable the detection of fish fraud at the protein level in the future.**

Guest edited as a special report on “Mass Spectrometry: Status Quo in Food Allergen and Food Authenticity Applications” by Bert Popping and Carmen Diaz-Amigo.
The authors declare not to have any conflict of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Color images are available online at http://aoac.publisher.ingentaconnect.com/content/aoac/jaoac
[1]Corresponding author’s e-mail: swittke@hs-bremerhaven.de
DOI: https://doi.org/10.5740/jaoacint.19-0061

The globalization of the flow of goods is causing a constant increase in the amount of fish sold or processed by the German fishing industry. It is estimated that more than 500 species of fish are available in Germany, of which approximately 87% are imported (1). It should be noted that about half of the imported goods come from non-European Union (EU) member countries (1), so these products are subject to the legislation of these countries. Nonetheless, these goods have to fulfill European regulations such as Regulation (EU) No. 1379/2013 on the common organization of the markets in agricultural products, which states that fish and fish products must be marked with the following information: (*1*) the commercial destination and the scientific name of the species, (*2*) an indication of the origin (trapped or bred), and (*3*) an indication of the geographical origin.

Despite these legal requirements, violations regularly occur. Recent research in the United States on the red snapper (*Lutjanus campechanus*), a high-priced reef fish native to the Gulf of Mexico, revealed that about 70–80% of the red snapper tested in the trade are declared wrongly. To maximize profits or to conceal illegally caught fish, for example, much cheaper and more widespread species are used, such as the redfish (*Sebastes* spp.) or the tilapia (*Oreochromis* spp.; 3–7). Although the fish itself can be easily morphologically identified, determining the species of (already processed) fish fillets is significantly more difficult and prone to errors. In this case, the raw materials must be examined and authenticated using laboratory analytical techniques. This also applies to distinguishing flatfish (Pleuronectiformes). Often, species such as dab (*Limanda limanda*) or Pangasiidae such as pangasius (*Pangasianodon hypophthalmus*) are advertised as high-priced species such as the European plaice (*Pleuronectes platessa*), sole (*Solea solea*), or turbot (*Scophthalmus maximus*), respectively (8). In addition, incorrect labeling will provide incorrect information with regard to the volume of a specific species caught.

For these reasons, the development of a method for the rapid and secure authentication of fish species or fish products prior to their placement on the market is very interesting for the processing and importing industries as well as for federal authorities.

In the past, species identification in food was often linked to a morphological analysis. Because of similar phenotypic

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

traits in seafood, this approach relatively quickly came to its limits, especially in already processed foods (9). As a result, DNA- (PCR) and protein-based methods (electrophoretic or immunological, such as ELISA) were established (7, 10, 11). For PCR-based methods, mixed samples and processed samples are commonly problematic because of highly degraded, nonamplifiable DNA or DNA that is somewhat unsuitable for the quantification of exact percentages of different species (12–15).

Although the genome in general does not change significantly throughout a lifetime, the proteome is dynamic. It does not change only because of its dependence on activated genes but also as a result of posttranslational modifications or external factors, such as the level of development of an organism, environmental changes, or interaction with other organisms.

Because of its good resolution and low-tech equipment, two-dimensional (2-D) gel electrophoresis is the most widely used method in proteomic research. Nevertheless, gel-based methods are reaching their limits. Hydrophobic and slightly soluble proteins such as membrane and basic proteins are not easily detectable by gel electrophoresis. In addition to difficulties in quantification, more highly concentrated proteins make it difficult to detect less abundant, and potentially differentiating, proteins (16). Cells consist of several tens of thousands of different proteins, which can differ enormously in their concentration and, therefore, impact the result of the 2-D gel electrophoresis. This makes it difficult to identify clear markers for species differentiation.

Overall, proteomes are different, less in the presence or absence of certain proteins than in the degree of excretion. Therefore, highly reliable and precise methods, e.g., for researching transmitter proteins in biological material, are necessary (16). In 1984, isoelectric focusing was accepted by AOAC INTERNATIONAL as the only validated protein-based method for distinguishing species (9).

For certain challenges (e.g., when examining highly processed or mixed samples), MS might be a better choice because of its higher resolution (17). MS, or more precisely, LC–MS, is routinely used for the discovery of species-specific peptides in reference samples and the detection of peptides (biomarkers) for diagnostic purposes in clinical proteomic research (18–20).

LC–MS and LC–MS/MS-based assays (in general, bottom-up proteomic assays) can be automated to deliver fast, reproducible results and provide a comprehensive diagnosis of samples as a result of high-throughput analysis. Untargeted identification and characterization of species-specific peptide markers would be the first step in developing fast and cost-effective methods, such as multiple reaction monitoring (MRM)–MS detection methods (7). MRM is a targeted method that offers a highly specific and simultaneous fragmentation of several target compounds (21).

Even if the technological basis is set, there are numerous problems with the use of mass spectrometric (LC–MS) methods. These include, for example, the choice of enzyme for proteolytic digestion. In most cases, trypsin is used, but other enzymes (such as thermolysin or the endoproteases Glu-C, Lys-C, etc.) might be suitable as well. In addition, parameters such as specificity, reproducibility, repeatability, etc., must be validated. Guideline Q2(R1) of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use provides a good overview of the parameters to be validated in the establishment of analytical methods (22).

The aim of the present work is, therefore, to investigate to what extent it is possible to establish an MS-based authenticity test of fish species by determination of peptide markers. First, it describes how the LC–MS/MSMS method was developed, standardized, and validated in-house. It subsequently describes, as a proof of concept, its ability to differentiate red snapper (*L. malabaricus*) from redfish (*Sebastes* spp.) and its specificity by using 68 samples from nine additional marine species as blinded controls.

To prove the feasibility of the method, two different approaches were used for the differentiation and statistical evaluation. On one hand, principal component analysis (PCA) was performed, and, on the other hand, artificial neural networks (ANNs), which can be particularly advantageous for complex data sets (23), were deployed.

Materials and Methods

Sample Preparation

For the optimization of the workup, all samples were lyophilized (Alpha 1-2 LDplus; Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany) and subsequently stored at –32°C until further use. All sample preparations were performed as duplicates and respective measurements as triplicates.

Thermolysin Digest (Final Method)

The enzymatic digestion of 0.3–0.5 g sample with thermolysin was carried out analogously to a protocol of Yokohama et al. (24). The lyophilized sample was dissolved in 8 mL ultra-pure water (Labtower™ 30 EDI; Thermo Fisher Scientific, Dreieich, Germany) and homogenized (Ultra-Turrax, 15 000 rpm, 1 min). After the proteins were exposed, the samples were heat-denatured (100°C, 10 min). Subsequently, after cooling (40°C), 3 mg thermolysin (*Geobacillus stearothermiphilus*; Sigma-Aldrich, Hamburg, Germany) was added and incubated (5 h, 40°C). Finally, the enzyme was deactivated by heat denaturation (100°C, 10 min).

Trypsin Digest

Tryptic digestion of the sample was carried out according to a guideline on peptide mapping by Agilent Technologies (25) with slight modifications in sample size and the denaturing process. To ensure pH stability, the sample was placed in 8 mL ammonium bicarbonate (100 mM; Carl Roth GmbH, Karlsruhe, Germany). After the proteins were exposed by Ultra-Turrax homogenization (15 000 rpm, 1 min), the samples were heat-denatured (100°C, 10 min). Subsequently, after cooling (60°C) and the addition of a further 4 mL ammonium bicarbonate, 200 µL dithiothreitol (DTT, 0.2 mol/L; Carl Roth GmbH) was added, followed by protein reduction (60 min, 60°C). Afterward, the alkylation (60 min, darkness) of thiol-groups was carried out by the addition of 800 µL iodacetamide (IAA, 200 mM/L; SERVA Electrophoresis GmbH, Heidelberg, Germany).

In the following step, the excessive IAA was separately deactivated by 200 µL DTT (0.2 mol/L, 60 min, darkness). After checking the pH (range 7–9), 3 mg trypsin (3 mg/mL trypsin in ammonium bicarbonate buffer; bovine pancreas, lyophilized, TPCK-treated, ≥10 000 units/mg protein; Sigma-Aldrich) was added and incubated (18 h, 37°C, darkness). Finally, the inactivation of excessive trypsin was affected by the addition of 200 µL formic acid (pH <4; Fluka, Seelze, Germany).

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

Protein Quantification (Final Method)

The photometric determination of the protein content was not performed initially and was only part of the final method. For this standardization, the UV absorption at 280 nm was used (Spectronic™ GENESYS™ 2; Thermo Fisher Scientific). Bovine serum albumin (Carl Roth GmbH) served as the calibration standard. The protein content of the samples was determined, and the protein concentration of all samples was finally adjusted to 2.0 mg/mL.

LC–MS and LC–MS/MS

LC–MS and LC–MS/MS analysis was performed using an Eksigent 200 MicroLC System online coupled to an AB Sciex TripleTOF® 4600 (AB Sciex, Darmstadt, Germany) in positive electrospray mode. Samples were separated on a HALO Fused-Core® C18 LC column (50 × 0.5 mm, 2.7 µm, 90 A; AB Sciex) at 35°C using a solvent gradient [running buffer A (LMA): water with 0.1% formic acid; running buffer B (LMB): acetonitrile with 0.1% formic acid] as described in Table 1. The flow rate was held at 20 µL/min, and an aliquot of 5 µL sample was injected. The mass calibration of the time-of-flight–MS was carried out for all four measurements with ß-galactosidase (0.25 pmol; AB Sciex) in TurbolonSpray® mode with a customized LC gradient (data not shown). In addition, signal intensities for typical MS/MS fragments of ß-galactosidase were monitored daily to ensure constant electrospray ionization conditions.

Measurements were carried out as an untargeted analysis in information-dependent acquisition (IDA) mode. If a signal exceeds a value of 100 counts/s (cps) and the ions are at least 2-fold charged, MS/MS spectra were recorded.

MS/MS peptide spectra were scanned within the UniProt protein database (version 2018_2) using the Paragon search engine (26). The database search using Protein Pilot™ (version 5.0) was performed using carbamidomethylation of cysteine as a fixed modification for samples after tryptic digest and without any fixed modification for all other samples (thermolysin digest). The MS and MS/MS mass tolerance was set to ±0.05 Da. All MS and MS/MS parameters are summarized in Table 2.

Statistical Analysis

For the visualization and definition of potential species-specific biomarkers, MarkerView™ 1.2 (AB Sciex) was used. In addition, for the comparison and classification of samples, a PCA after *t*-test was performed (MarkerView 1.2; AB Sciex). For QC purposes, all potential biomarkers were checked for accuracy and suitability on the basis of extracted ion chromatograms (XIC) and fragment spectra (PeakView™ 2.2; AB Sciex). The settings are shown in Table 3.

Table 1. Applied LC gradient

Time, min	LMA, %	LMB, %
0	95	5
0.5	95	5
21.5	65	35
22.0	10	90
27.5	10	90
28.0	95	5
32.0	95	5

Table 2. Ion source and MS/MSMS parameters

Parameter	Value MS/MSMS
ESI source, V	5200
Source temperature, °C	400
Curtain gas pressure, kPa	20
Collision energy, %	10
Nebulizer gas, kPa	15
Heating gas, kPa	25
Start mass, *m/z*	280/50
End mass, *m/z*	1250/1800

Normalization of LC–MS Data (Final Method)

To increase the comparability of the samples, the 100% method (MarkerView 1.2; AB Sciex) was used. The sum of all peak areas corresponds to 100%, and the proportions of individual peak areas are determined. In addition, a retention time (RT) correction was performed in MarkerView. For this purpose, a peak list was created, containing peaks that (*1*) were present in each sample (species-independent), (*2*) show a signal intensity above the limit of determination, (*3*) show no or very little dependence on the group, and (*4*) cover the entire separation in their entirety. Using this peak list, the correction was calculated as a function of the RT.

PCA and ANNs

PCA and ANNs were used to evaluate whether it is possible to distinguish between the two fish species on the basis of the complete LC–MS/MSMS data.

PCA simplifies multivariate data while losing as little information as possible. It allows a convenient comparison of the variability along the axis with the highest variability. In most cases, only the two axes with the greatest variability are retained, and the result can be displayed as a 2-D scatter plot (23).

Whereas PCA is an exploratory method for the identification of the most apparent differences in the data set, feed-forward ANNs can be used to classify fish species. ANNs require training data sets to train the neural network for the identification of fish

Table 3. Settings during data import and peak clustering in MarkerView™

Parameter	Value
Mass tolerance, Da	±0.1
Minimal RT, min	0.3
Maximal RT, min	23
RT tolerance, min	±0.5
Minimal spectral peak width, ppm	8
Minimal peak width (RT; scans)	4
Noise limit, cps[a]	100
Maximal number of peaks	15 000

[a] cps = Counts/s.

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

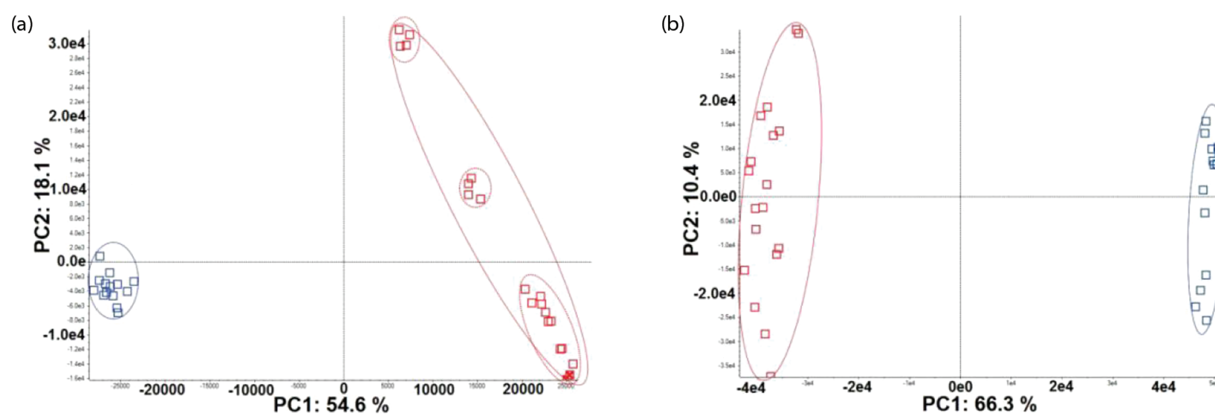

Figure 1. PCA plot for the differentiation of red snapper (n = 7; blue circles) and redfish (n = 10; red circles) samples after (a) thermolytic or (b) tryptic digest, respectively.

species. The result of the neural network is a classification score that represents the likelihood that the sample belongs to the respective species (23).

Results

Selection of Enzyme for Bottom-Up Approach

At the beginning of the method development, it was necessary to decide which enzyme was suitable for the authenticity testing of fish. Therefore, on the one hand, a standard protocol for thermolysin was used for the sample preparation, analogous to the procedure of Yokohama et al. (24). On the other hand, a standard protocol for protein digestion by trypsin (25) was used because trypsin is often the enzyme of choice in bottom-up proteomics studies.

Both sample preparations were applied to samples of red snapper (n = 7) and redfish (n = 10) and were examined by the same LC–MS and LC–MS/MS method. It became obvious that in tryptic digestion, a much higher number of peptides can be detected, and these peptides generally also show higher signal intensity.

In a second step, a selected red snapper sample was examined in more detail. Overall, more peptides were annotated by peak clustering in a database (MarkerView 1.2; Table 3) after tryptic digestion (4707 to 3689). In addition, more of these peptides were sequenced (1889 to 1431) in IDA mode, but fewer proteins matched existing records in the UniProt database (125 to 189). In addition, more high molecular weight peptides with molecular masses >4 kDa were detected (44 to 27) after tryptic digestion.

Finally, a PCA of red snapper versus redfish was executed to evaluate which sample preparation is better suited to differentiate the proteomes of these two species. In both cases (tryptic digest and thermolysin digest), a distinction between both groups was possible, but it was noticeable that no clusters had formed for redfish samples after tryptic digestion (Figure 1b), Unlike after thermolysin digestion (Figure 1a), the samples scattered similarly.

Based on this data, it was decided that further investigations would be carried out with thermolytic digestion. Crucially, because both approaches seem to be suitable to investigate fish fraud, the simpler sample preparation (no DTT, no alkylation) using thermolysin should facilitate transfer to other laboratories.

Furthermore, the PCA plot (Figure 1) indicates that, in the case of thermolysin digestion, the comparability of the samples examined is higher (especially shown for red snapper).

Optimization of the Thermolytic Digest

Thermolysin is a thermophilic enzyme with an optimum temperature of 60–70°C; nonetheless, sample preparation was initially carried out at 40°C. To evaluate whether higher temperatures improve the digestion, fish fillet samples of red snapper, redfish, salmon, and turbot were digested at five different temperatures (40, 50, 60, 70, and 80°C). As a marker peptide for the efficiency of digestion, Leu-Lys-ProAsn-Met (LKPNM), an angiotensin converting enzyme (ACE)-inhibiting peptide described by Yokohama et al. (24) that was detected in bonito after thermolytic digest, was used. In all species, the highest content of LKPNM was obtained at 40°C (blue), as exemplified for redfish in Figure 2. In addition, the comparability of the different measurements was very high, as could be demonstrated by PCA (data not shown).

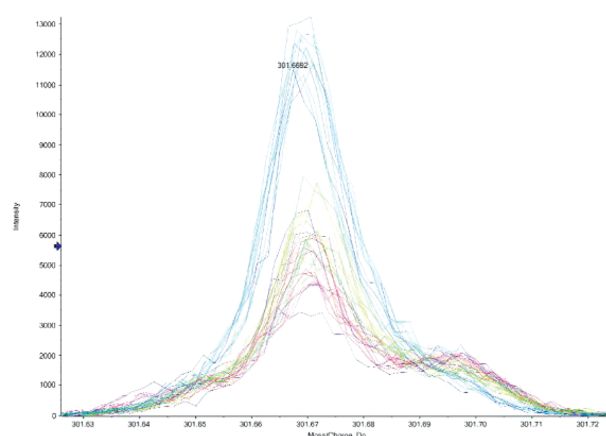

Figure 2. LKPNM content in redfish fillet after thermolytic digest. Values for 40, 50, 60, 70, and 80°C were plotted. All investigations were performed as triplicates in sample preparation (on three different days) and measurement (total n = 9). The blue lines (highest values) represent the nine LC–MS investigations for 40°C.

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

Furthermore, Figure 2 emphasizes the high reproducibility of the sample preparation because the sample preparation as well as the LC–MS/MSMS measurements were performed on three consecutive days.

Standardization of LC–MS (Evaluation of In-House Precision and Repeatability)

To further improve the comparability of LC–MS runs, several standardizations have been introduced. First, the aim was to make the signal intensities of repetitive experiments as comparable as possible. For this purpose, a standardization of the injected amount of peptide was performed. Based on protein quantification at 280 nm, each sample was diluted to a final protein concentration of 2.0 mg/mL. Furthermore, an intensity normalization using the total area sum method (MarkerView 1.2) was implemented.

Repeatability

Figures 3–5 illustrate that highly reproducible MS data were obtained after the execution of all standardization procedures, exemplarily shown for red snapper and pangasius. The red snapper data in Figure 3 represent nine LC–MS spectra for one red snapper fillet prepared by one operator in triplicate and measured three times.

Figure 4 represents similar investigations for one pangasius fillet and six red snapper fillets, respectively. The pangasius fillet (Figure 4a) was prepared twice and subsequently measured in triplicate. As a validation factor, the sample weight was varied by 20% (0.27 and 0.33 g). For red snapper (Figure 4b), the validation factor was to evaluate whether different samples prepared and measured on four different days by one operator were comparable.

It became obvious that the standardization of the protein quantity injected on the column as well as the intensity normalization were successful and led to a significantly improved reproducibility of the LC–MS assay.

To further assess the long-term repeatability of the assay, previously mentioned LC–MS and LC–MS/MS measurements of red snapper samples were annotated in a database. For this purpose, two measurement series of two different operators (series 1 and 2) were compared; with 12 months between the preparations, a sample aging process during storage might alter the results. A total of 10 preparations (three of one sample in series 1 and seven different samples for series 2) were compared with a total of 16 measurements from a total of seven different red snapper samples.

Despite the aging of the sample, following database annotation in MarkerView, it was found that, after applying the standardizations, a total of 6887 clusters were created. A total of 6461 signals (clusters) were detectable in all 16 measurements

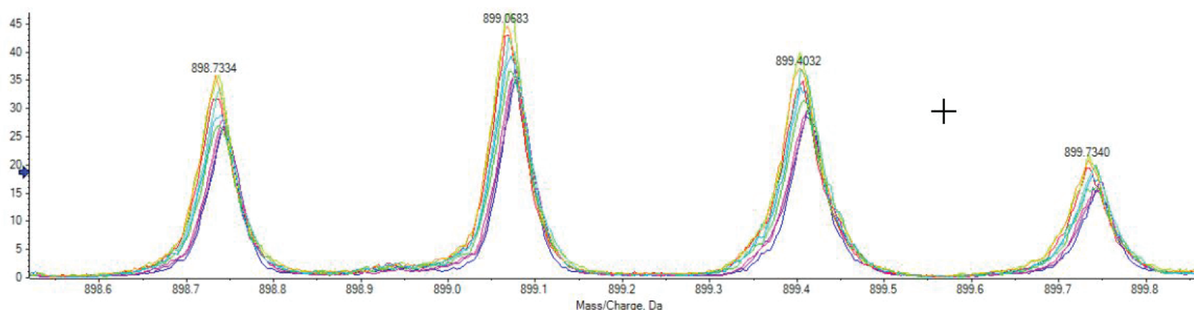

Figure 3. MS spectra in the range of 898.5–899.9 *m/z* for red snapper (*n* = 9, one sample; triplicates in preparation and measurement, one operator).

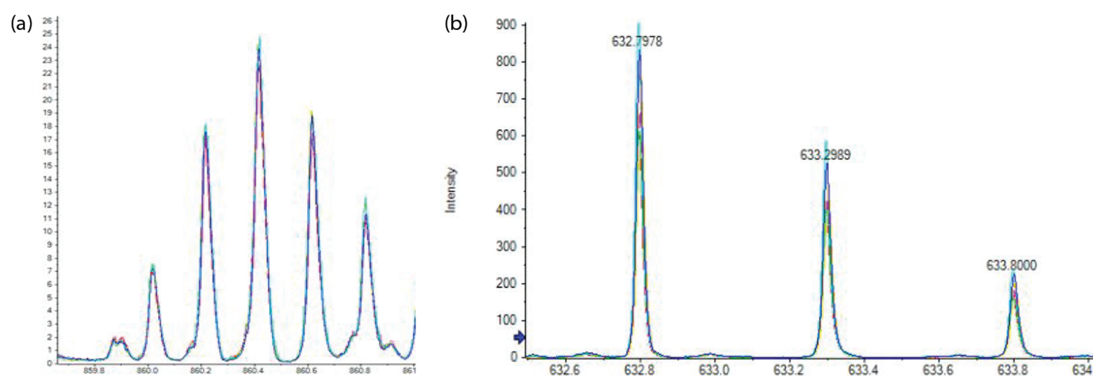

Figure 4. (a) MS spectra of a 5-fold charged peptide in the range of 859–862 *m/z* for one pangasius sample that was investigated in duplicate (three measurements each, one operator) using two different sample weights. (b) MS spectra for a 2-fold charged peptide in the range of 632.5–634 *m/z* for six red snapper fillets that were prepared on four different days (one measurement shown for each sample, one operator).

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

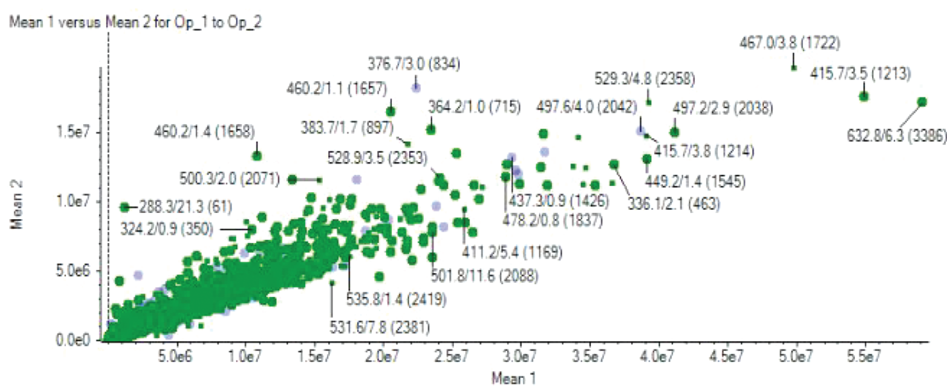

Figure 5. Comparison of mean values of two series conducted by two operators.

(94%) and only 111 (1.5%) in less than 10 measurements. Plotting the mean intensity values for both operators for all 6887 clusters resulted in a linear correlation, meaning that, in general, higher signal intensities in preparations of one operator correlate with increasing signal intensities of the other group (Figure 5).

In-House Intermediate Precision

Preparations and measurements of each of the two series were conducted under in-house intermediate conditions, i.e., not in parallel or subsequently but, in most cases, on different days. Series 1 measurements were obtained with the same individual fish, whereas variation within series 2 was inflated by several individual fish (of the same species). Thus, the variability within series 1 can be used to estimate the in-house reproducibility SD of the method between different days, whereas the SDs within series 2 could be greater than the in-house reproducibility SD. A display of both SDs by the corresponding mean value across the two series (the average of the geometric mean of series 1 and 2, respectively) is provided in Figure 6 as well as the mean relationship between the relative SD and peak size.

These calculations confirm that the average SD within series 2 is somewhat larger than the average in-house reproducibility SD obtained within series 1. However, this holds true only for relatively large peaks with a size of 104 and more, whereas for smaller peaks, the variation between individual fish is masked by larger in-house reproducibility SD analytical error. An overview of the estimated SDs is given in Table 4.

According to these in-house intermediate SDs, the average LOQ is around a peak size of 105. In other words, all peaks above this intensity value are promising candidates for species identification. This condition is fulfilled by more than 1200 peaks.

It should be noted that all these considerations were derived from the mean relationship between the relative SD and mean peak size. Actual SDs for individual peaks can be smaller or larger.

How Can the Differences Between Mean Values of the Two Series Be Explained?

Variation within the series under in-house intermediate conditions can partly explain the observed differences between the mean values of series 1 and 2, which were obtained by different operators. For 23% of all peaks, there is no significant difference between series 1 and series 2 (at significance level 5%). If one assumes that long-term variation with different operators is twice the variation within the series, for 48% of peaks there are no significant differences. This means that almost half of the peaks could be used for species differentiation, even if the samples are more than 1 year old.

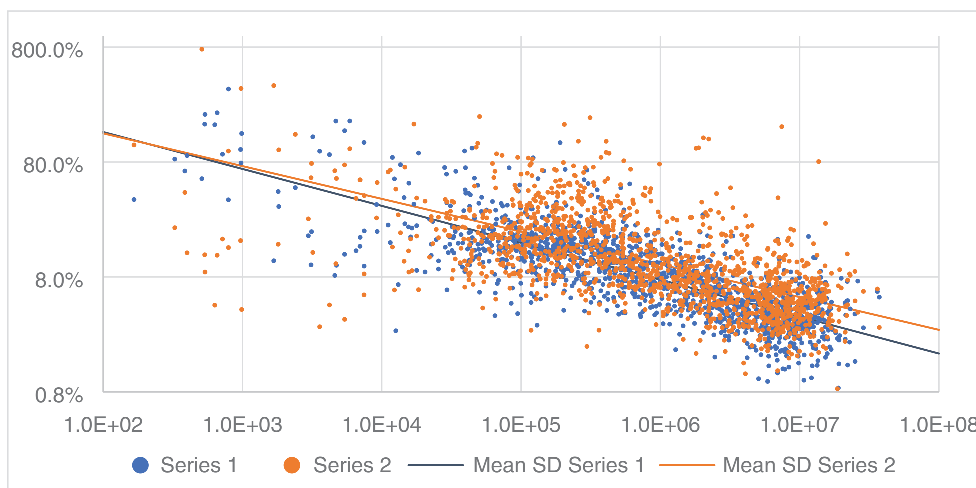

Figure 6. Relative in-house reproducibility SD for series 1 and series 2 (%).

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

Table 4. Comparison of the estimated SD for two measurement series addressing reproducibility

Size of peak	Average relative in-house RSD series 1, %[a,b]	Average RSD series 2, %[c]
1.0E+04	33.2	38.2
1.0E+05	15.9	19.8
1.0E+06	7.6	10.3
1.0E+07	3.6	5.3
1.0E+08	1.7	2.8

[a] RSD = Reproducibility SD.

[b] Series 1 describes the preparation of one red snapper sample (triplicate in preparation and measurement; $n = 9$).

[c] Series 2 describes the preparation of seven individual red snapper samples (monoplicates in preparation and measurement; $n = 7$).

Normalization of RT

Finally, because the RT slightly fluctuated in the LC–MS studies, a normalization of the RT was established based on a common peptide list (Table 5). This optimization led to the ability to improve the comparability of the RT of individual components over the entire LC–MS run. Originally, the time deviation for identical signals was in the range of about ±2 min (Figure 7a) and could be reduced to <±0.5 min with the help of this approach (Figure 7b). Because the definition of biomarkers requires data synchronization in a database, a smaller time deviation helps to make clustering in the database less error prone.

Differentiation of Fish Species

After a reproducible in-house method was established for determining the proteome of fillet samples of different fish species, this approach was exemplarily applied to the differentiation of red snapper and redfish because red snapper is regularly replaced by the less expensive and more common redfish (3–6). In a first effort, PCA (Figure 1, thermolysin digest) and ANNs were applied to the complete data set (MarkerView; Table 3) to evaluate the feasibility of the assay.

According to Figure 1, PCA allows a distinct differentiation between the two groups. In addition, a feed-forward ANN was used to differentiate among three fish classes through a multiclass classification: red snapper, redfish (fish2), and "Miscellaneous," which represents 68 measurements of nine

Table 5. List of 34 *m/z* values used for retention time normalization

m/z, Da	RT, min	Charge state	*m/z*, Da	RT, min	Charge state
280.1856	2.44	1	542.7576	14.93	2
283.6685	5.29	2	546.2956	5.24	2
344.8596	0.80	1	563.3159	2.83	1
371.9162	4.05	3	574.6035	5.93	3
384.6964	0.8	2	580.2993	5.83	2
407.2704	3.47	2	625.9612	4.13	3
428.8943	2.45	4	655.7781	5.94	2
431.2559	1.35	2	659.3488	9.10	2
439.8863	7.09	3	669.8579	5.62	2
449.5035	8.27	3	684.3133	8.83	2
454.587	6.12	3	708.4593	9.54	1
490.2506	0.87	2	721.0431	10.95	3
507.2677	6.11	1	787.404	7.99	2
510.623	6.69	3	796.4104	6.52	2
524.0886	9.34	5	848.333	2.07	1
535.2327	5.81	2	862.4064	7.88	2

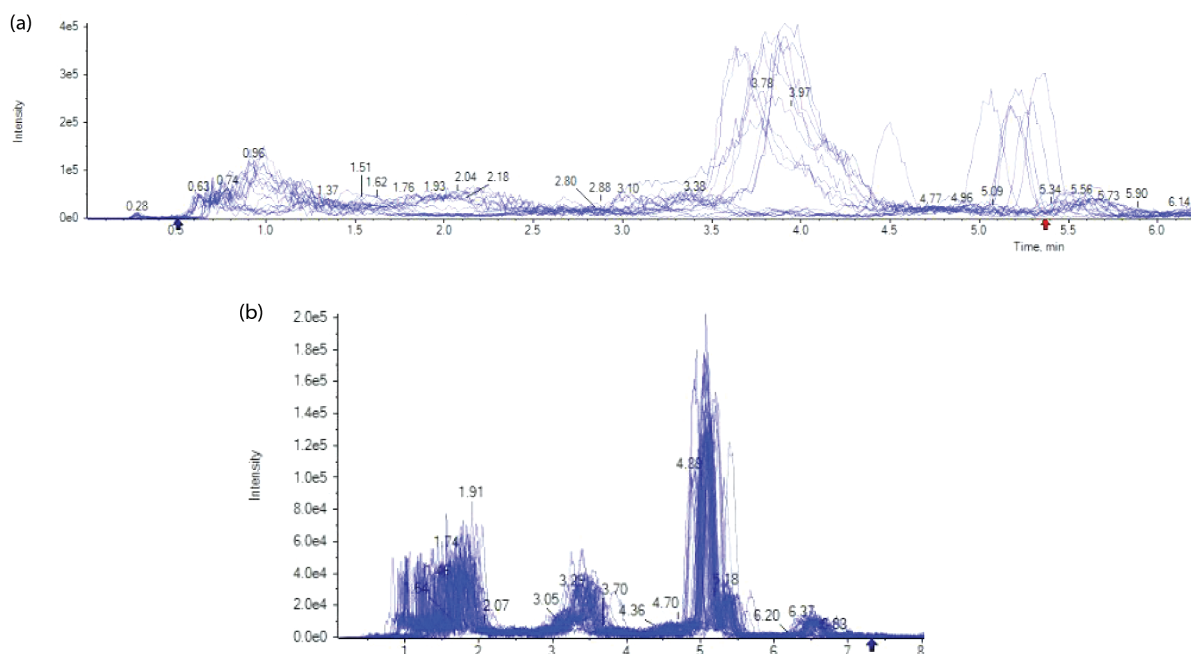

Figure 7. Reduction of the time deviation for individual *m/z* signals (exemplarily shown for a signal with *m/z* = 412.8; 0.5–6 min) by applying a retention time standardization based on internal standards. (a) XIC ahead of RT normalization and (b) XIC after RT normalization.

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

Table 6. ANN classification scores

	n	Minimum	Maximum	Mean	SD
Red snapper	14	1.93	4.83	3.91	0.95
Fish2[a]	15	–3.91	–2.80	–3.59	0.27
Misc[b]	68	–11.56	–5.46	–9.36	1.09

[a] Fish2 = Redfish.

[b] Misc = Nine different marine species.

additional marine species such as pangasius (*P. hypophthalmus*), salmon (*S. salar*), turbot (*S. maximus*), plaice (*P. platessa*), sole (*S. solea*), lemon sole (*Glyptocephalus cynoglossus*), halibut (*Reinhardtius hypoglossoides*), red salmon (*Oncorhynchus nerka*), and great scallop (*Pecten jacobaeus*) as negative controls. The classification score presented in Table 6 represents the likelihood that the sample is indeed a red snapper. The higher the classification score, the higher the probability that the sample is indeed a red snapper. According to these results, there is no red snapper sample for which the classification score is negative or close to zero. For all other samples, the classification score is clearly negative, suggesting that the ANN was able to differentiate clearly between red snapper and other fish species. This means also that species-specific characteristics of the spectrum could be reproduced successfully by the method, so that finally the ANN could classify the species correctly (Table 6).

Calculations of the ANN were conducted by means of nested cross-validation to avoid overinterpreting the data set (27); i.e., the data set was split into five folds so that, for each fold, the classification score was computed on the basis of the other four folds.

This untargeted approach is highly effective but would bind too much on measurement and bioinformatics capacity and, in addition, compromise platform independence because of the poorer detection limit and specificity compared with a targeted MRM assay. To address this issue, it is therefore necessary to identify and validate species-specific biomarkers. Moreover, in this procedure, more and larger groups (in this study, species and samples per group) are to be examined in comparison, which then can serve as a negative control or to define further questions (28). Because the focus of this publication is on the description of method development, the in-house validation, and its feasibility, the definition of biomarkers will be exemplarily shown for the differentiation of red snapper and redfish.

The first step is to define selection rules for biomarkers, with the following list containing initial suggestions: (*1*) potential biomarkers must have a frequency of occurrence of 100% in at least one of the groups, irrespective of the pretreatment and storage of the samples (fresh, frozen, or aquaculture); (*2*) signal intensities must have a sufficiently high S/N, as signals are only accepted above the LOW (1:10 as S/N); (*3*) only monoisotopic signals are allowed as biomarkers; (*4*) signals which are only 1-fold charged ($z = 1$) are excluded as potential marker peptides if no corresponding 2-fold charged signal exists; (*5*) potential markers that could be traced back to the same protein after MS/MS examination and database comparison may only be included once in the biomarker list (to avoid bias); and (*6*) to ensure that no signals with similar *m*/*z* coelute, each potential marker peptide must be checked by means of an XIC and/or uniquely identified by the MS/MS spectrum.

Applying these selection rules resulted in the definition of potential species-specific biomarkers, allowing a differentiation between red snapper and redfish. One of these potential biomarkers, detectable in red snapper but not in redfish, is shown in Figure 8.

Conclusions

In summary, we have succeeded in developing a method for distinguishing fish species that has proven to be reproducible as an in-house assay. The method has been applied to a total of 16 species (in addition to bonito, tuna, oyster, and blue mussel) and has also proven to be successful in this comparison (Wittke, S., University of Applied Sciences Bremerhaven, unpublished data, 2019). Thus, a promising method has been developed that could enable the detection of fish fraud at the protein level in the future.

To evaluate the method externally, an interlaboratory test has now been initiated that will be carried out in three stages. In the first, currently ongoing study, LC–MS platforms are to be identified that have a sufficiently good detection limit to be usable for the research question in the future. The second stage then includes performing the complete assay (sample preparation,

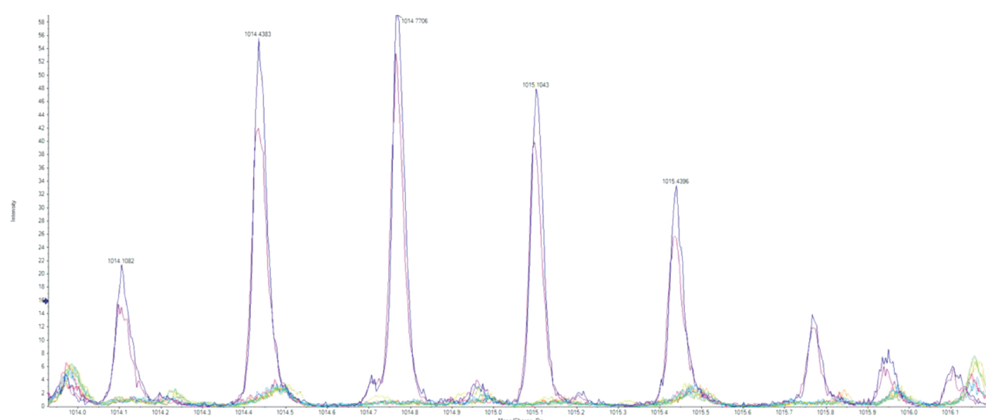

Figure 8. Potential biomarker for red snapper. The peptide (*m/z* 1014.1802, 3-fold charged, 3039.32 Da) can only be detected in the red snapper samples, as exemplarily shown for two red snapper and five redfish samples.

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024

measurement, and evaluation) in external laboratories. In the third part, it is then necessary to validate the species-specific biomarkers in a blinded study using an MRM approach.

References

(1) *Fisch-Informationszentrum e. V. Fischwirtschaft, Daten und Fakten 2017 im Überblick* (2017) https://www.fischinfo.de/images/broschueren/pdf/FIZ_DF_2017.pdf (accessed on October 27, 2018)

(2) The European Parliament and the Council of the European Union, Regulation (EU) No. 1379/2013, https://eur-lex.europa.eu/eli/reg/2013/1379/oj (accessed on October 27, 2018)

(3) Verrengia, J.B. (2004) Today's Catch: Red Snapper Fakes, http://www.cbsnews.com/news/todays-catch-red-snapper-fakes/ (accessed on October 27, 2018)

(4) Hsieh, Y-H. P., Woodward, B.B., & Blanco, A.W. (1994) *J. Food Qual*. 18, 131–140. doi:10.1111/j.1745-4557.1995.tb00368.x

(5) Marko, P.B., Lee, S.C., Rice, A.M., Gramling, J.M., Fitzhenry, T.M., McAllister, J.S., Harper, G.R., & Moran, A.L. (2004) *Nature*, 430, 309–310. doi:10.1038/430309b

(6) U.S. Department of Health and Human Services, Regulatory Fish Encyclopedia, Seafood Species Substitution and Economic Fraud, http://www.fda.gov/Food/FoodScienceResearch/RFE/ucm071528.htm (accessed on October 27, 2018)

(7) Rasmussen, R.S., & Morrissey, M.T. (2008) *Comp. Rev. Food Sci. Food Saf.* 7, 280–295. doi:10.1111/j.1541-4337.2008.00046.x

(8) Pardo, M.A., Jimenez, E., & Perez-Villareal, B. (2016) *Food Control* **62**, 277–283. doi:10.1016/j.foodcont.2015.10.048

(9) A. Cifuentes (Ed.) (2013) *Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition*, Wiley, New York, NY

(10) Lockley, A.K., & Bardsley, R.G. (2000) *Trends Food Sci. Technol*. **11**, 67–77. doi:10.1016/S0924-2244(00)00049-2

(11) Rehbein, H.Z. (1990) *Z. Lebensm.-Unters. Forsch*. **191**, 1–10. doi:10.1007/BF01202356

(12) Hellberg, R.S.R., & Morrissey, M.T. (2011) *J. Lab. Autom*. **16**, 308–321. doi:10.1016/j.jala.2010.07.004

(13) Farag, M.R., Alagawany, M., Lagawany, M.E., El-Hack, A., Tiwari, R., & Dhama, K. (2015) *Adv. Anim. Vet. Sci*. **3**, 334–346. doi:10.14737/journal.aavs/2015/3.6.334.346

(14) Koh, M.C., Lim, C.H., Chua, S.B., Chew, S.T., & Phang, S.T.W. (1998) *Meat Sci*. **48**, 275–285. doi:10.1016/S0309-1740(97)00104-6

(15) Zhang, J., & Chai, Z. (2006) *Eur. Food Res. Technol.* 223, 413–417

(16) Pischetsrieder, M., & Baeuerlein, R. (2009) *Chem. Soc. Rev.* **38**, 2600–2608. doi:10.1039/B817898B

(17) Nessen, M.A., Van de Zwaan, D.J., Grevers, S., Dalebout, H., Staats, M., Kok, M., & Palmblad, M. (2016) *J. Agric. Food Chem.* **64**, 3669–3677. doi:10.1021/acs.jafc.5b05322

(18) Lopez, J.L., Marina, A., Alvarez, G., & Vazquez, J. (2002) *Proteomics* **2**, 1658–1665. doi:10.1002/1615-9861(200212)2:12<1658::AID-PROT1658>3.0.CO;2-4

(19) Wittke, S., Mischak, H., Walden, M., Kolch, W., Rädler, T., & Wiedemann, K. (2005) *Electrophoresis* **26**, 1476–1487. doi:10.1002/elps.200410140

(20) Carrera, M., Canas, B., Pinero, C., Vazquez, J., & Gallardo, J.M. (2007) *J. Proteome Res.* **6**, 3070–3080. doi:10.1021/pr0701963

(21) Gross, J.H. (2013) *Massenspektrometrie: Ein Lehrbuch*, Springer Spektrum Verlag, Heidelberg, Germany

(22) International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (2005), Validation of Analytical Procedures: Text and Methodology Q2(R1), https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf (accessed on March 14, 2019)

(23) Uhlig, S., Colson, B., Hettwer, K., Simon, K., Uhlig, C., Wittke, S., Stoyke, M., & Gowik, P. (2018) Valid machine learning algorithms for multiparametermethods, Eurachem 2018, Dublin, Ireland, https://eurachem.org/images/stories/workshops/2018_05_Data/pdf/5_Bertrand_Colson_Valid_machine_learning_algorithms.pdf (accessed on March 14, 2019).

(24) Yokoyama, K., Chiba, H., & Yoshikawa, M. (1992) *Biosci. Biotechnol. Biochem.* **56**, 1541–1545. doi:10.1271/bbb.56.1541

(25) Agilent Technologies, Schlüsselstrategien zur optimalen Characterisierung von Peptiden: Ein praktischer Leitfaden für Peptid-Mapping, https://www.agilent.com/cs/library/selectionguide/public/5991-2348DEE.pdf (accessed on March 14, 2019)

(26) Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M., & Schaeffer, D.A. (2007) *Mol. Cell Proteomics* **6**, 1638–1655. doi:10.1074/mcp.T600050-MCP200

(27) Liotta, L.A., & Petricoin, E.F. (2010) *Clin. Chem.* **56**, 1641–1642. doi:10.1373/clinchem.2010.146142

(28) Mischak, H., Apweiler, R., Banks, R.E., Conaway, M., Coon, J., Dominiczak, A., Ehrich, J.H., Fliser, D., Girolami, M., Hermjakob, H., Hochstrasser, D., Jankowski, J., Julian, B.A., Kolch, W., Massy, Z.A., Neusuess, C., Novak, J., Peter, K., Rossing, K., Schanstra, J., Semmes, O.J., Theodorescu, D., Thongboonkerd, V., Weissinger, E.M., Van Eyk, J.E., & Yamamoto, T. (2007) *Proteomics Clin. Appl*. **1**, 148–156. doi:10.1002/prca.200600771

Downloaded from https://academic.oup.com/jaoac/article/102/5/1330/5658264 by guest on 20 April 2024