The Use of Peptide Markers of Carp and Herring Allergens as an Example of Detection of Sequenced and Non-Sequenced Proteins

Justyna Bucholska* and Piotr Minkiewicz
University of Warmia and Mazury in Olsztyn, Faculty of Food Sciences, Department of Food Biochemistry, Plac Cieszyński 1, PL-10-726 Olsztyn, Poland

Received: May 6, 2015
Accepted: March 23, 2016

Summary
The objective of this study is to identify fish protein markers for detecting multiple species based on a comparative proteomic approach that relies on fragments with identical sequences. The possibilities and challenges of the use of peptides obtained from carp (Cic- rinus carpio) and herring (Clupea harengus) proteins are discussed. A bioinformatic analysis was followed by an LC-MS/MS experiment to identify markers predicting the presence of fish allergenic proteins. Selected myosin peptides were found in carp protein hydrolysates with known sequences and in herring protein hydrolysates with unknown sequences. The results obtained for carp and herring proteins myosin and parvalbumin indicate that proteins with unknown sequences can be identified by peptide markers. Such markers can be designed by disregarding the principle that peptides should be unique (present in one sequence). The challenge is to determine a group of proteins that can be detected by peptide identification.

Key words: allergens, proteins, fish, markers, peptides, mass spectrometry

Introduction
Food allergies, including allergies to fish and fish proteins, pose a significant challenge for food science and medicine (1,2). Food scientists continue to search for new analytical methods that effectively detect allergens in foods. Therefore, it is important to develop analytical methods for the identification of allergens. Mass spectrometry is one of the methods recommended for this purpose (3–5).

Peptides belong to a group of compounds that provide information about the composition of food ingredients and food products (6–9). The analyzed peptides are fragments of allergenic proteins or proteins originating from organisms that are known sources of allergens (3–5). On-line liquid chromatography with mass spectrometry is an effective and popular tool for analyses of allergenic proteins. Trypsin (EC 3.4.21.4), an enzyme frequently used in protein hydrolysis (3,8–10), and pepsin (EC 3.4.23.1), a proteolytic enzyme commonly used in studies of allergenicity (10), are used to identify proteins (11).

Johnson et al. (4) formulated several recommendations concerning the selection of peptides to be used as allergen markers. One of the main principles they proposed is the uniqueness of peptides. A peptide should be unique for precursor proteins. Parvalbumins or their fragments are recommended as unique markers of species-specific fish allergens due to high sequence variability (8,12). A single peptide marker capable of detecting several allergens can also be identified (13). A group of allergens that can be identified by a single peptide belongs to the same family in the AllFam database (14). According to the principles of comparative proteomics, a change in the paradigm determining the choice of peptide biomarkers could support the identification of proteins with un-
known sequences (15,16). The carp (Cyprinus carpio) is an example of a fish species with many known protein sequences, whereas the herring (Clupea harengus) has never been studied extensively with the aim of determining its protein sequences. The UniProt database (17) screening using the carp Latin name resulted in 2564 protein sequences, whereas the herring Latin name occurs only in 275 items (status from 07.03.2016). In both species the number of sequences includes items considered as putative or derived from homology, but not found at protein level.

In Allergome (18), the biggest database of allergens, the listed allergens are not only proteins, but also tissues or species, e.g. fish. For example, the herring (Clupea harengus, Allergome code 1370) and the carp (Cyprinus carpio, Allergome code 1979) are listed together with different proteins found in these species. In this database, allergens can be detected by identifying protein fragments synthesized by the analyzed organism. Fragments of allergenic proteins do not have to be always detected. In addition to parvalbumins, the best known fish allergens (2), other proteins may also be used as biomarker precursors. Myosins are the most prevalent myofibrillar proteins, and their fragments are generally easy to detect. Sequence similarity was reported among various myosins (19), which makes them good candidates for comparative proteomic analysis. Carp myosins are broadly represented in the UniProt database (17), whereas very few sequences of herring myosins are known.

The aim of this study is to identify fish protein markers for detecting multiple species based on a comparative proteomic approach that relies on fragments with identical sequences. The possibilities and challenges of the use of peptides obtained from myosin and parvalbumin fish proteins are discussed.

Materials and Methods

In silico analysis

Sequences of 112 carp (Cyprinus carpio) proteins and 19 herring (Clupea harengus) proteins from the UniProt database (17) were used. Most of the analyzed fish proteins were myosins and parvalbumins. Parvalbumins are considered as an example of peptides adequate for use as unique markers, whereas myosins, due to sequence conservation, as precursors of peptides sufficient for analysis with use of comparative proteomic principles.

The first stage of the study was the in silico analysis of carp and herring protein sequences to identify fragments of allergenic proteins that differ from fragments of non-allergenic proteins. The analysis was performed using EVALLER software (20).

The program identified three fragments of individual proteins that best met the above criterion. Similarities with known fragments of allergenic proteins were identified based on the Smith-Waterman score (21). Protein fragments were submitted for further analysis when the score was 84 or higher (minimum for proteins considered by the program as allergens).

The next stage of the study involved in silico proteolysis of potentially allergenic proteins selected by the EVALLER program. In silico proteolysis was performed using the PeptideMass application (22). Two enzymes, trypsin (EC 3.4.21.4) and pepsin (EC 3.4.23.1), were used to identify fragments released after proteolysis. The fragments produced by in silico proteolysis were regarded as potential markers if they were part of the fragments displayed in EVALLER. Peptides containing at least seven amino acid residues were submitted for further analysis. Fragments with a mass-to-charge ratio higher than 2000 Da were excluded. If more fragments produced by in silico proteolysis were displayed in one protein sequence, one fragment from carp protein and two fragments from herring protein with the highest sequence cross-coverage (SCC) between the proteolytic fragment and the fragment displayed in EVALLER were selected. SCC was calculated (and expressed in percentage) with the use of the following equation (13):

\[ SCC = \frac{N_e}{N_e + N_p} \times 100 \]

where SCC is the sequence cross-coverage between the expected proteolysis product and the corresponding epitope, \(N_e\) is the number of amino acid residues in the expected proteolysis product and in the fragment predicted by the EVALLER program, \(N_p\) is the number of amino acid residues in the sequence of the fragment predicted by the EVALLER program, and \(N_e\) is the number of amino acid residues in the sequence of the predicted proteolytic product.

The SSRCalc program (23), with a correction proposed by Dziuba et al. (24), was used to predict peptide retention times. The correction was introduced using the following equation:

\[ t_{predicted} = -0.0562t_{predicted}^2 + 1.7428t_{predicted} + 6.7354 \]

where \(t_{predicted}\) is the predicted retention time calculated in the SSRCalc program, and \(t_{corrected}\) is the predicted retention time calculated based on the correction proposed by Dziuba et al. (24).

The following parameters were used to calculate the \(t_{predicted}\): elution time of unretained compounds 2.02 min, parameter \(b\ 0.94\), and pore diameter (closest to pore diameter in the applied column) 100 Å (24).

MS/MS spectra were simulated in the fragment ion calculator program (25). Various types of typical fragment ions, including neutral loss (release of water or ammonia) (26), were taken into account.

Peptides from protein sequences listed in the UniProt database (17) were identified experimentally with the use of the previously described (27) parameters (PAM10 matrix, expected threshold 1000) in the WU-BLAST program (28). The preceding amino acid residues that formed bonds susceptible to hydrolysis by trypsin or pepsin were also included, excluding the N-terminal parvalbumin fragment, according to previous recommendations (13).

Experimental analysis

Carp and herring were purchased from a local market (Olsztyń, Poland). Carp was supplied by the Szwaderki fish farm (Olsztyń, Poland), and herring was harvested from the Baltic Sea. Carp and herring were pur-
chased as fresh carcasses and transported immediately to the laboratory where they were filleted, portioned and packaged on the same day. The prepared experimental material was frozen at –70 °C.

Sarcoplasmic proteins were extracted from both fish species by the methods proposed by Bugajska-Schretter et al. (29) and Carrera et al. (12). Extraction of sarcoplasmic proteins was carried out using frozen fish. A mass of 100 g of carp or herring white muscle was homogenized with two volumes of 10 mM Tris-HCl, pH=7.2, supplemented with 5 mM phenylmethylsulfonyl fluoride (PMSF), using a commercial blender (Waring, Torrington, CT, USA) for approx. 1 min. Fish extracts were centrifuged (centrifuge model 3K30; Sigma Laborzentrifugen GmbH, Osterode am Harz, Germany) at 40 000×g for 1 min at 4 °C. The supernatants were filtered using membrane filters (0.22 μm; Whatman, GE Healthcare Life Sciences, Dassel, Germany), lyophilized and stored at −70 °C until analysis.

Myofibrillar proteins were isolated according to the method described by Martinez et al. (30). Before extraction, carp or herring fillets were placed in a freezer at −20 °C. White fish muscle was scraped while still frozen. A volume of 80 mL of Tris, pH=10.5, was added to 6 g of white fish muscle and homogenized for approx. 1 min. After homogenization, samples were centrifuged at 15 000×g for 7 min at 4 °C. After centrifugation, the supernatants (Tris extracts) were collected and frozen at −70 °C. A volume of 40 mL of the solution of 8 M urea, 4 % 3-(3-cholamidopropyl)dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPS), 2 mM tributyl phosphate (TBP), 40 mM Tris and 0.2 % IPG (immobilized pH gradient) were added to the pellet. Samples were homogenized for 30 s, centrifuged at 15 000×g for 7 min at 4 °C, and the supernatants (CHAPS-urea extracts) were frozen at −70 °C in the pellet. Samples were lyophilized and stored at −70 °C. Fish proteins were extracted from the same fish in triplicate.

The protein content of lyophilisates was determined according to the method proposed by Bradford (31). The analyses were performed in triplicate.

Sarcoplasmic and myofibrillar proteins were hydrolyzed after the extraction from fish specimens. Specific proteolysis was performed using two proteolytic enzymes: bovine pancreatic trypsin, catalog number T1426, and porcine pancreatic pepsin, catalog number P7012, both from Sigma-Aldrich, St. Louis, MO, USA.

Hydrolysis was performed under the following conditions: protein concentration 3 mg/mL, enzyme concentration 150 μg/mL, pH=8.0 for trypsin and 2.0 for pepsin, temperature 37 °C, and time of hydrolysis 24 h. The enzymatic reaction was stopped by deactivating the enzyme at a temperature of 100 °C for 5 min (32). Immediately after hydrolysis, the samples were frozen at −70 °C and lyophilized.

The solutions containing 0.5 mg/mL of the fish protein isolate soluble in salt solutions and their hydrolysates (trypsin and pepsin) were analyzed. The samples were dissolved in 6 M urea solution in a mixture of acetonitrile and water at a ratio of 100:900 by volume, pH=2.2, with the addition of trifluoroacetic acid (TFA) according to the method described by Visser et al. (33). A Shimadzu (Tokyo, Japan) set comprising two LC-10AD pumps, an SCL-10AD autosampler, an SCL-10AD controller, a CTO-
-10AS thermostat and an SPD-M10AW photodiode detector with a Jupiter Proteo Phenomenex® (Torrance, CA, USA) column, 250 mm×2 mm, particle diameter 4 μm, and pore diameter 90 Å, was used in the analysis. The Class-VP 5.0 Shimadzu® application was used for data analysis. Solvent A was 0.01 % (by volume) TFA solution in water. Solvent B was 0.01 % (by volume) TFA solution in acetonitrile. The gradient of solvent B was increased from 0 to 40 % during 60 min. The column was washed (40–100 % B for 60 to 65 min, 100 % B for 65 to 70 min) and equilibrated (100–0 % B for 70 to 71 min, and 0 % B for 71 to 80 min) (24,34). Data acquisition time was 80 min, flow rate 0.2 mL/min, injection volume 10 μL and column temperature 30 °C.

The RP-HPLC-MS/MS analysis was performed in a Varian 500-MS (Agilent Technologies, Santa Clara, CA, USA) ion trap mass spectrometer with electrospray ionization connected to an HPLC assembly containing two 212-L pumps, ProStar 410 autosampler, Degasit degasser (MetaChem Technologies®, Torrance, CA, USA) and 2-2 nitrogen generator (Parker Domnick Hunter Scientific®, Gateshead, UK). The column, solvent system, gradient and other HPLC separation parameters were identical to those described above. Data acquisition time was 5–60 min. Mass spectrometry parameter were as follows: nee- dle and shield voltages: 5000 and 600 V respectively, spraying and drying gas (nitrogen) pressure 35 and 30 psi respectively, drying gas temperature 390 °C. The remaining parameters: positive polarity, capillary voltage 100 V, re- tardation factor loading 100 %, isolation window 3.0, excita- tion storage level m/z=206.3, excitation amplitude 2.98– 3.28 V, syringe volume 250 μL, sample loop volume 100 μL, needle tubing volume 15 μL, flush volume 100 μL, col- umn oven setpoint 30 °C, frequency data recording 0.05–0.07 Hz, single scan averaged from five microscans, op- tions such as: use of air segment, headspace pressure and alarm buzzer were included (24,34). Retention times of peptides were determined after smoothing the al- gorithm described by Savitzky and Golay (35) as recom- mended previously (24). RP-HPLC-MS/MS analyses of hydrolysates were performed in duplicate.

Results and Discussion

The results obtained with the use of three protein se- quences: herring and carp parvalbumins and carp myosin heavy chain are presented in Fig. 1. The complete se- quence of herring myosin is not available in the UniProt database (17). To date (07.03.2016) only two myosin frag- ments, with accession numbers Q98ST0 and Q90ZP0 can be found in the UniProt database (17) using ‘Clupea harengus’ together with ‘myosin’ as a query. Their length covers less than one tenth of myosin heavy chain sequence. The sequences of herring myosin fragments do not contain subsequences selected in silico as potential protein mark- ers, although they are very similar to carp myosin se- quences. The list of peptide sequences identified using the in silico analysis is summarized in Table 1. Parvalbu- mins form a group of proteins with high sequence vari- ability, therefore, a peptide that is present in more protein sequences represented by that family is difficult to detect. The SCC value of this peptide is presented in Table 1, and
it is attributed to carp parvalbumin. All fragments generated by proteolysis simulation belong to the fragments identified in the EVALLER program (20). The value of the SCC (13), discussed in this work, is the ratio of the length of the fragment from the proteolysis simulation to the length of the fragment identified in the EVALLER program (20). The EVALLER program was originally designed to predict protein allergenicity (20). In this study, this application was used only to select possible protein fragments characterized by the highest similarity to the fragments of allergenic proteins. Protein fragments displayed by the EVALLER program often overlap with known sequential epitopes (36).

RP-HPLC was used to monitor proteolysis. Chromatograms of particular carp and herring protein fractions and products of their hydrolysis by pepsin are presented in Fig. 2. Intact proteins from isolates of sarcoplasmic and myofibrillar proteins were the dominant fractions with retention times exceeding 70 min (Figs. 2a, c and e). These fractions disappeared during proteolysis (Figs. 2b, d and f). The dominant protein hydrolysate fractions were eluted in 20 to 70 min. In these chromatograms, relative peak area between 10 and 70 min ranged from 90 to 99% of total peak area with retention times exceeding 10 min. Peaks eluted before 10 min contain unretained substances, such as components of buffers for dissolving proteins or peptides (33). Similar results were obtained using hydrolysis by trypsin.

Identical gradients and columns were used for RP-HPLC and RP-HPLC-MS/MS analyses. The difference in the time of analysis and retention times of particular fractions resulted from variations in dead volume of both HPLC assemblies.

LC-MS/MS chromatograms of the DKKNVIRL peptide from carp and herring myofibrillar protein hydrolysate are shown in Fig. 3. Peptides identified by LC-MS/MS are listed in Table 1. The results of in silico analysis: a) herring (Clupea harengus) parvalbumin sequence (accession number C6GKU6 in the UniProt database (17), Allergen Clu h 1.0101 (18), Allergen code 6101 (18)), b) carp (Cyprinus carpio) parvalbumin sequence (accession number P02618 in the UniProt database (17), Allergen Cyp c 1 (16), Allergen code 263 (18)), and c) carp (Cyprinus carpio) myosin sequence (accession number Q90339 in the UniProt database (17)). Fragments displayed by the EVALLER program (20) are underlined, fragments resulting from simulated proteolysis are indicated in italics.

Table 1. Peptides from carp and herring proteins selected in silico as potential markers

<table>
<thead>
<tr>
<th>Peptide sequence</th>
<th>Precursor*</th>
<th>SCC max/%</th>
<th>M/Da</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFLQNFSAGAR</td>
<td>parvalbumin: P09227 (carp)</td>
<td>39.3</td>
<td>1222.6</td>
</tr>
<tr>
<td>AFAVGNLADIAAELACK</td>
<td>parvalbumin: P02618 (carp)</td>
<td>79.2</td>
<td>1861.9</td>
</tr>
<tr>
<td>LFLQNFK</td>
<td>parvalbumin: P02618 (carp); C6GKU6 (herring)</td>
<td>25.0</td>
<td>908.5</td>
</tr>
<tr>
<td>MAFAGVLNDADITAELACK</td>
<td>parvalbumin: Q8ULS2 (carp)</td>
<td>80.0</td>
<td>2023.0</td>
</tr>
<tr>
<td>GADIDAILK</td>
<td>parvalbumin: C6GKU6 (herring)</td>
<td>22.0</td>
<td>872.5</td>
</tr>
<tr>
<td>ALTDAETK</td>
<td>parvalbumin: C6GKU6 (herring)</td>
<td>33.3</td>
<td>847.4</td>
</tr>
<tr>
<td>EDAITAGALK</td>
<td>parvalbumin: C6GKU8 (herring)</td>
<td>48.0</td>
<td>1161.6</td>
</tr>
<tr>
<td>MAFAAFLK</td>
<td>parvalbumin: C6GKU8 (herring)</td>
<td>32.0</td>
<td>897.5</td>
</tr>
<tr>
<td>QAAEEAEQTNHLSR</td>
<td>myosin: Q90339 (carp)</td>
<td>25.9</td>
<td>1714.8</td>
</tr>
<tr>
<td>VQLHAQANTSLNQK</td>
<td>myosin: Q2HX56 (carp)</td>
<td>27.3</td>
<td>1705.9</td>
</tr>
<tr>
<td>AAEAEAEQANSNLTK</td>
<td>myosin: Q76FW4 (carp)</td>
<td>50.0</td>
<td>1603.7</td>
</tr>
<tr>
<td>ADIAESEQVK</td>
<td>myosin : Q76FW6 (carp)</td>
<td>52.6</td>
<td>1073.5</td>
</tr>
</tbody>
</table>

*Accession numbers in the UniProt database are given. SCC=sequence cross-coverage
lysates are presented in Fig. 3 (fragment ions named according to Roepstorff and Fohlman (37)). Peptides are displayed as groups of fragment ions detected at the same retention time (34,38–40). Peptide fragmentation involved the formation of various types of fragment ions, including products of neutral loss. A list of experimentally identified peptides is presented in Table 2. Only 10 out of 15 peptides selected in silico were identified experimentally. Peptides were regarded as identified if they formed a group of fragment ions with identical retention times. In line with the previous recommendation (24), the difference between predicted and measured retention times should not exceed 10%. The risk of unsuccessful identification was discussed in a previous study (34). Unsuccessful identification could be attributed to the absence of fragmentation in an ion trap mass spectrometer or differences in retention time from the predicted value. The applied protocol determines which peptides can be identified (41). Set of peptides possible to be identified (so-called proteotypic peptides) may vary due to mass spectrometer type (e.g. matrix-assisted laser desorption ionization vs. electrospray). None of the methods guarantees identification of all possible products of proteolysis (41). Proteins may undergo modification when the molecular mass of protein fragments changes or when flanking bonds become resistant to proteolysis. Such modifications were responsible for the fact that some peptides, selected in silico as carp myosin fragments, were detected only in hydrolysates of herring myofibrillar proteins (Table 2). Sequences of herring myosins remain unknown, but they contain fragments identical to those of carp myosins. This suggests that potential precursors of the peptides listed in Table 2 could involve many more proteins with unknown sequences.
are smoothed with the Savitzky-Golay (35) algorithm. Dziuba et al. (24) observed that this algorithm does not always support the generation of unambiguous retention times.

The results of BLAST search are presented in Table 2. Proteins containing peptide sequences whose flanking bonds are susceptible to trypsin or pepsin were divided into three categories. The first category covers parvalbumins and myosins from both edible and non-edible fish. The second category includes proteins from edible animals. Edible species are those used in the food industry, such as chicken (Gallus gallus) and turkey (Meleagris gallopavo). This category does not include some exotic animals or animals that are consumed incidentally in certain countries. In this study, animals of the type are regarded as inedible and are placed in the third category. According to the recommendations formulated by Johnson et al. (4), peptides belonging to the first and third category are potential allergen markers.

Johnson et al. (4) proposed a set of criteria for selecting peptide markers. They are: known sequence of precursor proteins, uniqueness, absence of chemical or enzymatic modifications in the peptide sequence, possibility of protein extraction from food ingredients or food products, possibility of peptide release with the use of selected proteolytic enzymes, and proteolysis-resistant proteases in the peptide sequence. Excluding the last two criteria, the above requirements have to be fulfilled for peptide detection. Modifications of amino acid residues can change the molecular mass of peptides or make the preceding and successive bonds inaccessible for proteolytic enzymes used in hydrolysis (trypsin or pepsin in this study). Non-extractable proteins are not available for proteolysis. The predicted proteolytic pattern is achieved when the last two criteria are met. The last two criteria were confirmed by the detection of peptides. The peptides that do not meet any one of the last two criteria could not be identified in this study.

The choice of peptides that do not originate from proteins with known sequences and are not unique for one protein creates new possibilities. The number of protein sequences in databases such as UniProt (17) is rapidly growing, but many sequences remain unknown. Protein sequences, including sequences of allergenic proteins, have not been studied extensively in all fish species. The above justifies the use of the principle of comparative proteomics which states that identical or highly similar fragments may occur in homologous proteins with known and unknown sequences (15). Our findings indicate that identical fragments released from carp and herring myosins can be used as protein markers. In our previous study (13), the same fragment (containing at least seven amino acid residues) was detected in homologous proteins belonging to the same family, identified based on the presence of an appropriate domain. Allergenic proteins that belong to the same family in the AllFam database (14) are usually characterized by cross-reactivity. Allergens of carp and herring belong to the same families of proteins and their presence may be detected via identification.

### Table 2. Peptide markers of carp and herring proteins identified experimentally

<table>
<thead>
<tr>
<th>Peptide sequence</th>
<th>Carp protein hydrolysate</th>
<th>Herring protein hydrolysate</th>
<th>m/z Da</th>
<th>t&lt;sub&gt;retained&lt;/sub&gt; min</th>
<th>t&lt;sub&gt;sequential&lt;/sub&gt; min</th>
<th>Protein detected based on the identified peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFLQNFK</td>
<td>sarcoplasmic proteins</td>
<td>sarcoplasmic proteins</td>
<td>453.3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>39.1</td>
<td>36.0–36.2</td>
<td>39 fish parvalbumins, 14 proteins from inedible organisms</td>
</tr>
<tr>
<td>MAFAGVLDADITAALECK</td>
<td>sarcoplasmic proteins</td>
<td></td>
<td>1012.5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>54.5</td>
<td>56.7–57.7</td>
<td>1 fish parvalbumin</td>
</tr>
<tr>
<td>AAEAEAEEQANSNLTK</td>
<td>myofibrillar proteins</td>
<td></td>
<td>802.9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>21.7</td>
<td>22.7–23.0</td>
<td>1 fish myosin</td>
</tr>
<tr>
<td>DKKNINRL</td>
<td>myofibrillar proteins</td>
<td></td>
<td>500.8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>18.5</td>
<td>20.0–20.9</td>
<td>16 fish myosins</td>
</tr>
<tr>
<td>INTKKKL&lt;sup&gt;c&lt;/sup&gt;</td>
<td>myofibrillar proteins</td>
<td></td>
<td>844.6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>17.1</td>
<td>17.8–18.2</td>
<td>58 fish myosins, 64 proteins from other edible animals, 169 proteins from other organisms</td>
</tr>
<tr>
<td>DKKNVIRL</td>
<td>myofibrillar proteins</td>
<td>myofibrillar proteins</td>
<td>985.6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>22.9</td>
<td>19.8–21.5</td>
<td>6 fish myosins, 1 protein from an organism not used in the food industry</td>
</tr>
<tr>
<td>ALTDAETK</td>
<td>sarcoplasmic proteins</td>
<td></td>
<td>848.4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>17.6</td>
<td>17.2–17.3</td>
<td>84 fish parvalbumins, 2 proteins from inedible organisms</td>
</tr>
<tr>
<td>GADIDAALK</td>
<td>sarcoplasmic proteins</td>
<td></td>
<td>873.5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>26.5</td>
<td>26.0–26.1</td>
<td>1 fish parvalbumin</td>
</tr>
<tr>
<td>ADIAESQVNK</td>
<td>myofibrillar proteins</td>
<td></td>
<td>537.8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20.8</td>
<td>18.6–19.2</td>
<td>195 fish myosins, 108 proteins from other edible animals, 197 proteins from other organisms</td>
</tr>
<tr>
<td>QAEAAEQTNTHLSR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>myofibrillar proteins</td>
<td></td>
<td>871.9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20.6</td>
<td>21.3–21.5</td>
<td>1 fish myosin</td>
</tr>
</tbody>
</table>

<sup>a</sup>m/z=(M+2H+)/2, <sup>b</sup>m/z=M+H+, <sup>c</sup>peptide does not fulfill the criteria proposed by Johnson et al. (4). It was included in the table to demonstrate the problems associated with the detection of protein groups based on a single peptide marker.
cation of the same markers. Peptides found within this experiment may serve as such markers.

The rapid increase in the number of known protein sequences also poses a significant problem. Unique peptides that are markers of individual proteins are increasingly difficult to find. A peptide that is initially regarded as unique may become a group marker when new protein sequences homologous to its first known precursor are discovered. New precursors can also be found in the evaluated groups of proteins and organisms. In our study, this problem was solved by using a peptide with flanking bonds susceptible to trypsin or pepsin. Present in proteins isolated from edible mammals, birds, reptiles or amphibians. According to a less restrictive version of the criterion proposed by Johnson et al. (4), a peptide from more than one precursor may be accepted as a marker if additional homologues are found in food components. This restriction effectively prevents a false positive result. In our study, one peptide was a marker of a group of fish allergens when it occurred only in fish proteins or proteins from inedible vertebrates.

Short peptides may be attributed to a family of homologous proteins based on the presence of an appropriate domain (13,15,16,17,42). Examples of peptides used as allergen markers and originating from more than one precursor were discussed in our previous publication (42). The examples provided in the above reference concern peptides from bovine milk and chicken egg proteins. Their precursor proteins (α-, casein and lysozyme C) reveal interspecies conservation, understood as the presence of common fragments. Fish myosins possess the same property. Peptides that are potential markers of a smaller group of proteins are difficult to find in analyses of single taxa such as fish. The search for an appropriate peptide set requires simulation of proteolysis. The number of the resulting fragments is then analyzed using BLAST (28) or a similar application. When protein sequences have been retrieved, additional information relating to the taxonomic lineage of species synthesizing these proteins has to be found. Lastly, evidence indicating that these species are suitable for human consumption has to be provided. Species that are presently not suitable for food production could be used by the food processing industry in the future. Further research is needed to devise a reproducible and rapid method of peptide selection. In this study, the EVALLER program (20) was used to speed up the process of peptide selection. This strategy delivered satisfactory results for parvalbumins, but it was less accurate for myosins. Peptide markers of the myosin tail family in the InterPro database (signature IPR002928) (43) are easy to find, but a limited group of markers specific for individual taxa within this family is more difficult to identify. In the AllPro database (14), the myosin tail family is listed as a family of allergenic proteins (AF100), but invertebrate myosins are allergens. Vertebrate proteins from the myosin tail family, including myosin heavy chains, are not considered to be allergens. The number of myosins that are single peptide precursors accounts for approx. 15% of the total number of proteins belonging to the myosin tail family (Table 2). There are no simple algorithms for defining groups of allergens that can be detected based on a single marker. Further research is needed to examine the prevalence of myosins of different animal organisms and their suitability for human consumption.

Although this study included fresh fish, mass spectrometry together with various separation techniques was applied to identify peptides from processed fish, e.g. cooked, canned, high-pressure-treated before freezing, as well as peptides from parasites in fillets (44–47).

The detection of proteins without known sequences on the basis of peptide identification has a weak point. Even absolute quantification of a peptide cannot give information about allergen content if we do not have additional information about the source of proteins. Content of proteins being precursors of marker peptides may vary among species. Peptides from parvalbumins may serve for identification of species and for quantification of proteins. Identification of peptides from myosins can help us to find proteins from myosin family attributed to fish and occurring together with fish allergens, such as parvalbumins. This information is incomplete. Even absolute quantification of a peptide cannot give information about allergen content if we do not have additional information about the source of proteins (content and ratio of particular proteins may vary among species). Peptide sequences attributed to protein families (defined according to InterPro database (43)) instead of single sequence do not allow species identification. Using unique peptides our alternative offers complete and precise information if we find a peptide originating from a precursor with known sequence or lack of information if we do not know such precursor. The presence of a peptide attributed to a protein family provides partial information, but allows warning about the presence of allergen even if there is no known sequence. Peptides attributed to protein families instead of single sequences, as recommended by Schevchenko et al. (15), may support allergen detection. Use of a more abundant peptide within a family may increase likelihood of detection of a protein belonging to this family, but not sequenced to date.

This work can be considered as a preliminary study concentrated on the opportunity to find peptides potentially serving for detection of allergic fish, with unknown sequences. Myosins, occurring together with allergenic proteins, are highly conserved and provide better opportunity to find unknown allergens than parvalbumins, possessing species-specific sequences. Many analytical problems are still to be solved, such as how to convert limits of detection of particular peptides into limits of detection of particular proteins, tissues and species, and how to quantify uncharacterized allergen on the basis of determined amount of peptide.

Conclusions

An analysis of carp and herring proteins confirmed the possibility of finding peptides that are markers of proteins with unknown sequences. Such markers can be designed by abandoning the principle that peptides should be unique (should occur in one sequence only). Parvalbumin fragments of the analyzed fish can be recommended as unique markers, while myosin fragments may be recommended as group markers. Peptide markers could be fragments of allergenic proteins or proteins that are present with them and derived from the same organism. In this experiment we identified ten peptide markers of...
and by the University of Warmia and Mazury in Olsztyn, and Higher Education as part of grant No. NN 312 286 33 9.

8. References

Con

Acknowledgements

This study was supported by the Ministry of Science and Higher Education as part of grant No. NN 312 286 33 9.

Conflict of Interest

The authors declare no conflict of interest.

References


30. Martinez I, Siżykte R, Daukšas E. High resolution two-dimensional electrophoresis as a tool to differentiate wild from farmed cod (Gadus morhua) and to assess the protein composition of klipfish. Food Chem. 2007;102:504–10. http://dx.doi.org/10.1016/j.foodchem.2006.03.037


