Olsztyn, dnia 23.05.2022

Allergens in food can be identified using peptides that are markers of their presence. Markers specific to a given family of proteins can be used to detect the presence of allergens of unknown amino acid sequences based on similarity of their sequences to the sequences of proteins from phylogenetically related species. Identification of peptides derived from proteins with unknown sequences is performed using reversed phase high performance liquid chromatography coupled with tandem mass spectrometry (RP-HPLC-MS / MS) technique.

The aim of this research was to identify peptides from tropomyosin in processed seafood products. In the first stage, a bioinformatic analysis of selected tropomyosin amino acid sequences was performed. In a laboratory experiment, proteins were extracted from 4 products that had been defrosted, cooked or fried without fat, as well as from two ready-to-eat products. Proteins were hydrolyzed with trypsin, freeze-dried and the hydrolysates were analyzed by RP-HPLC-MS / MS technique. The selected raw materials contained species both with known and unknown tropomyosin sequences.

By *in silico* analysis, 72 amino acid sequences of the peptides (7 to 15 amino acids long) were chosen. They were products of simulated trypsin proteolysis of at least two tropomyosin sequences. Then, they were analyzed for their presence in the total number of tropomyosin sequences from the UniProtKB database. The obtained result was converted into the percentage of this number in the total number of tropomyosins in the Interpro and Pfam databases of protein families.

After RP-HPLC-MS / MS analysis, 40 peptides present in at least two samples were selected. Then, those that were not among the products of simulated trypsin proteolysis (contained missed cleveages) were separated. These peptides (23) were analyzed for the presence of fragments of allergenic tropomyosin epitopes from the Immune Epitope Database. The remaining 17 peptides, products of simulated trypsin proteolysis, were also analyzed to see if they were epitope fragments. In both cases the sequence cross coverage (SCC) was calculated. To check which families of proteins from which a given peptide can be released (after digestion with trypsin), the Unipept tool was used, and then phylogenetic trees were generated, which illustrate the taxonomic similarity between the organisms being the sources of precursor proteins of the analyzed peptides. The distribution of these peptides in the sequences of proteins from the tropomyosin family was also analyzed (based on the results of the *in silico* analysis).

Using the PEAKS software, the degree of coverage of the tropomyosin amino acid sequences present in the sample with the tropomyosin amino acid sequences available in the

UniProtKB database was determined. The software identified the analyzed peptides as fragments of known allergenic tropomyosins. Cooking the raw material made it possible to obtain longer fragments after digestion with trypsin, which made it possible to obtain a higher degree of sequence coverage. Frying, on the other hand, due to changes in the protein structure, significantly reduced the possibility of covering the amino acid sequence with the use of computer algorithms. In this case it was necessary to use de novo sequencing.

The use of a universal peptide marker, characteristic of the protein family, enables the detection of an allergen without the need to know its amino acid sequence. This approach, proposed in this paper, can be considered a reversal of the traditional allergen detection paradigm, which involves searching for peptides unique to a given protein amino acid sequence.