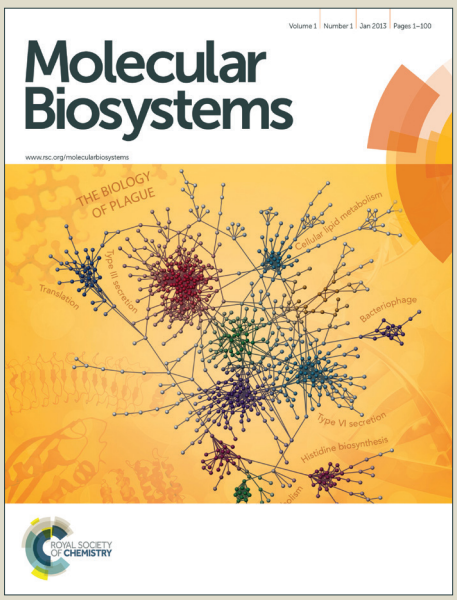
Molecular BioSystems

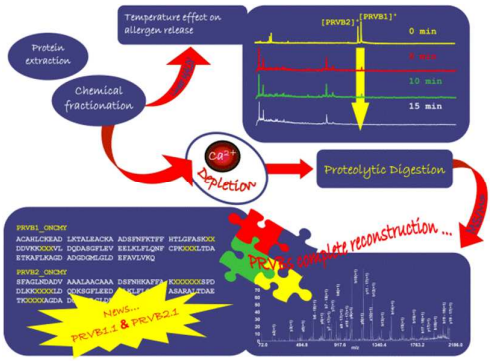
Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Graphical abstract



Journal Name

RSCPublishing

ARTICLE

Major allergen in rainbow trout (*Oncorhynchus mykiss*): complete sequences of Parvalbumin by MALDI tandem mass spectrometry

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Donatella Aiello[a], Stefano Materazzi[b], Roberta Risoluti[b], Hariprasad Thangavel[a], Leonardo Di Donna[a], Fabio Mazzotti[a], Francesca Casadonte[a], Carlo Siciliano[c], Giovanni Sindona[a] and Anna Napoli[a*]

Fish Parvalbumin (PRVB) is an abundant and stable protein in fish meat. The variation in cross-reactivity among individuals is well-known and explained by a broad repertoire of molecular forms and differences between IgE-binding epitopes in fish species. PVRB has "sequential" epitopes which keep their IgE-binding capacity and allergenicity also after heating and digestion with proteolytic enzymes. From the allergonomics perspective, PRVB is still a challenging target due to its multiple isoforms present at different degrees of distribution. Little information is available in the databases about PVRBs from *Oncorhynchus mykiss*. At present, only two validated, incomplete isoforms for this species are included in the protein databases: Parvalbumin Beta 1 (P86431) and Parvalbumin Beta 2 (P86432). A simple and rapid protocol has been developed for selective solubilization of PRVB from the muscle of farmed rainbow trout (*Oncorhynchus mykiss*), followed by calcium depletion, proteolytic digestion, MALDI MS, and MS/MS analysis. With this strategy thermal allergen release was assessed and PRVB1 (P86431), PRVB1.1, PRVB2 (P86432) and PRVB2.1 variants from the rainbow trout were sequenced. The correct ordering of peptide sequences was aided by mapping the overlapping enzymatic digests. The deduced peptide sequences were arranged and the theoretical molecular masses (Mr) for the resulting sequences were calculated. Experimental masses (Mr) for each PRVB variants were measured by Linear MALDI-TOF.

Introduction

Seafood plays an important role in human diet. The increasing consumption of fish in the everyday diet has enhanced the adverse reactions to fish allergens.[1-3] The immune-mediated systemic allergic reactions to fish are not only restricted to ingestion but include manual handling and inhalation of cooking vapours in the domestic and occupational environment.[4-6] Parvalbumin (PRVB) represents the major fish allergen which has been described in a wide range of fish species and belongs to the biggest group of animal-derived food allergens, the helix-loop-helix (EF-hand) domain family (http://www.meduniwien.ac.at/allergens/allfam/), with over 63 allergens currently reported. It is a calcium (Ca)-binding protein of low molecular weight (10-12 kDa) and has acidic isoelectric point (pI 3.0-5.0) with no disulfide bonds, and is characterized by three helix-loop-helix binding motifs, two helices pack together at an angle of ~90 degrees, separated by a loop region where calcium ion can be bound.[7] Only two of the three helix-loop-helix (EF-hand) motifs of parvalbumin are functional and can chelate metal ions, where the first domain provides a cap that covers the hydrophobic surface of other two domains.[8,9] PRVBs can be found in two distinct phylogenetic lineages: α and β, based on their amino acid composition.[10] α-PRVB consists of 109 aa with a less acidic pI (~5), and in general, it is not allergenic, while β–PRVB comprises 108 aa and has more acidic pI. Although fish contains both α and β parvalbumin, the majority of allergens reported in literature belongs to β lineage.[11] β–PRVBs, which are found in high amounts in the sarcoplasmic fraction of white muscle of fishes, are water-soluble proteins with remarkable heat stability.[12] The most part of immunoglobulin E (IgE) mediated reactions can remain even after heat treatment.[13,14] The allergenic properties of PRVBs are related with their resistance to gastrointestinal enzymes and their heat resistance, even if their IgE-binding capacity, and consequently the allergenicity, is reduced upon depletion of calcium-ions.[15] The detection of fish parvalbumin is challenging compared to other food allergens. This can be attributed to the high biochemical and immunological variability among the different fish species. Furthermore, it is known that some parvalbumin variants are able to generate oligomers after heating. The difference in thermal stability of PRVBs across different genera of fish indicates that heat processing has a strong effect on antibody reactivity of different parvalbumins. The PRVB from rainbow trout (Oncorhynchus mykiss) was selected as a case study, since it is one of the main fish species farmed in Europe and is found in two different colours of flesh, white and red. Trout farming is carried out for

various purposes and in a wide variety of production systems, however there has been no systematic survey of production systems and husbandry procedures at European level. Usually, salmonids which are not able to synthesize pigments are supplied with feeds containing Xanthophylls added in broodstock diet to improve egg quality or in on-growers diet to colour flesh before slaughtering. Xanthophylls can be found in natural matrices and added to the diet at a level of 20 % or synthetic Xanthophylls can be used and added to commercial feed.[16] Canthaxanthin (Cx) and Astaxanthin (Ax) are the most frequently added Xanthophylls to the feed. The fundamental ''Regulation (EC) No 1831/2003''(2003)[17] determines the use of additives in animal nutrition and sets out rules for the authorization, marketing, and labelling of feed additives. Ax /Cx are allowed in feeding salmon and trout (Commission regulation (EC) No1288/2004).[18,19] Their use is permitted only from the age of six months onwards and the mixture Ax/Cx is acceptable provided that the total concentration of astaxanthin and canthaxanthin does not exceed 50 mg/kg in the complete feedingstuff. Several speculations have been proposed about Ax and Cx binding sites,[20] and it is interesting that apparently no other teleost fishes are able to accumulate Ax/Cx in the muscles as salmonids.[21] Ax binding in muscle is through non-specific association of the myofibrillar proteins by means of weak hydrophobic interactions.[22] Xanthophylls are associated with hydrophobic areas of proteins, especially with the lipid component of lipoproteins, in order to be transported in an aqueous environment.[23]

From the allergonomics perspective, PRVB shows some interesting challenges due to its multiple isoforms present at different degree of distribution. Thus, an MS-based proteomics study of multiple forms of PRVB from whole antigen extract could constitute a significant contribution. Mass spectrometric methodologies provide a rapid and sensitive tool for the identification[24] and quantitation of metabolites,[25-27] amino acids,[28] and proteins[29,30] and their post-translational modifications.[31] Recent trends in proteomics suggest direct analysis of protein mixtures, with less emphasis on preliminary protein purification steps.[32] For these reasons, we have adopted simple chemical fractionation procedure of the whole proteome extracts from white and red muscle flesh to obtain samples containing only one family of proteins, and/or sets of chemically homogeneous proteins. Fractionation of the sample coupled with MALDI mass spectrometry represents the methodology of choice to get reliable results in the profiling of sarcoplasmic trout proteins. The PRVB enriched fractions were used to obtain the antigen structural characterization by tandem mass spectrometry and to test its thermal stability.

Results and discussion

A simple and rapid protocol has been developed for selective protein solubilization from white and red muscle flesh of farmed rainbow trout, followed by calcium depletion, proteolytic digestion, MALDI MS and MS/MS analysis. With this strategy thermal allergen release was assessed and PRVB variants from trout were sequenced.[33,34] The results of MS profiling and top-down proteomics approach from white and red rainbow trouts are described in the following sections.

Protein profile of trout extracts

PRVB is a homogenous acidic protein belonging to a group of muscle sarcoplasmic proteins. It is a thermo-stable marker protein for Isoelectric Focusing (IEF) gels, and is used for differentiation and identification of fish species even after heating.[35] Purified or enriched fraction of PRVB can be obtained from denaturing extraction of homogenized tissue followed by heating.[36,37] The use of saline solution containing sodium chloride (3-5%) or Tris-HCl (pH 7.2, 5 mM, PMFS) to prepare the sarcoplasmic extract are not amenable to direct MALDI-MS analysis. Furthermore, the disruption of all non-covalently bound protein complexes into individual polypeptides could compromise the objective of part of this study. Therefore, we decided to perform the proteome extraction in water to promote a selective protein solubilization (sarcoplasmic proteins) and fractionate the crude extract. The chemical fractionation was performed by adopting a saline solution (NH_4HCO_3, 50 mM) in presence of an organic co-solvent to promote the formation of carboxylate/ammonium ion pairs and hold the solubility of acid proteins (Scheme, experimental section). The protein content of the fractions $A_{r\text{-}w}$, $B_{r\text{-}w}$ and $C_{r\text{-}w}$ was monitored by linear MALDI MS. The data set displayed by the three spectra provides the entire profiling of sarcoplasmic proteins. The PRVB enriched sample (A_w) was determined by the appearance of the ion species of m/z 5640, 5886, 11280 and 11772, corresponding to doubly and mono-charged forms of the expected antigen, respectively (Figure 1A). The predicted MWs of the PRVB variants are 11345 Da and 11800 Da for Parvalbumin β2 and β1 (UniProt: PRVB2_ONCMY, P86432 and PRVB1_ONCMY, P86431), respectively. Therefore, it can confidently be suggested that the ion peaks of m/z 5886 and 11772 correspond to the PRVB variant β1 ($[PRVB1]^{2+}$ and $[PRVB1]^{+}$, Figure 1A), while those of m/z 5640 and 11280 can be ascribed to the PRVB variant β2 ($[PRVB2]^{2+}$ and $[PRVB2]^{+}$, Figure 1A). The MALDI spectrum of fraction B_w is characterized by the presence of the same protein profile (Figure 1ESI).

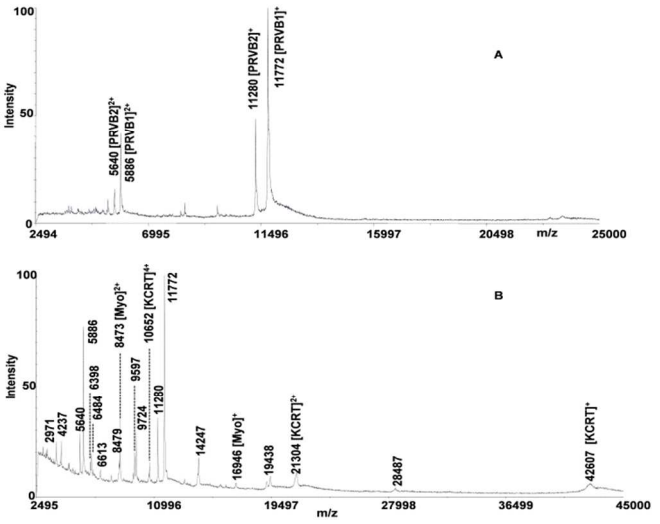


Figure 1. Linear MALDI MS spectra of (A) fraction A_w, (B) fraction C_w.

On the contrary, the MALDI spectrum of fraction C_w (Figure 1B) differs considerably from the previous spectra, since it is characterized by the appearance of several ionic species at higher molecular weight. These ions correspond to basic or neutral water-soluble proteins. According to the experimental conditions all fractions might contain sarcoplasmic proteins, many of which are enzymes associated with energy-producing metabolic pathways. Except for PRVBs, these enzymes have

only partially been characterized.[38] Therefore, the ion peaks at m/z 42607, 21304 and 10652 could be ascribed to $[KCRT]^{+}$, $[KCRT]^{2+}$ and $[KCRT]^{4+}$ of Creatine kinase (KCRT_ONCMY, P24722, Figure 1B). The ion species of m/z 8473, and 16946 can confidently be ascribed to doubly and mono-charged forms of myoglobin (Myo), respectively.[39] The data set displayed by the three MALDI spectra of red rainbow trout showed similar protein profiles (Figure 2ESI). Also in this case, the first fraction A_r contains only two PRVB variants. The fractions A_r and A_w can be taken into account for allergen release and structural characterization.

Effect of heat treatment.

The ability of PVRBs to withstand heat-treatment[13] can be attributed to its exceptionally stable secondary structure.[15] Nevertheless not much information is available on the effect of thermal processing for specific detection of PVRBs from white and red rainbow trout. Homogenised red and white rainbow trout fleshes were cooked by bain-marie mode at 100°C, for 0, 5, 10 and 15 minutes, then the samples were extracted and fractionated as reported in the Experimental Section. Figure 2-I displays the Linear-MALDI-MS spectrum of fraction A_w after 0, 5, 10, 15 minutes of cooking. MALDI MS spectra revealed that the cooking process has significantly modified the protein profiles. Heating of white flesh muscle has a profound effect on PVRB detection. The molecular peaks found at 11-12 kDa disappear suggesting that both variants of PRVB are completely thermally denaturized after 5 minutes of cooking in white rainbow trout. Extra experiments indicated that the complete denaturation of the same variants occurs after only 3 minutes of cooking (Figure 3ESI).

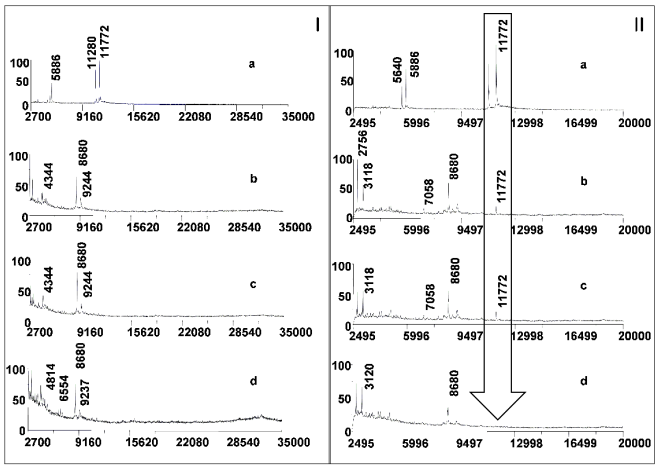


Figure 2. Linear MALDI MS spectra of fraction A after (a) 0 min, (b) 5 min, (c) 10 min, (d) 15 min of cooking (at 100°C) from homogenised white (panel I) and red (panel II) trout fleshes.

The lower detection of white PVRB variants after heating could be due to unfolding and/or aggregation processes.[40] The last is a complex phenomenon that can involve other compounds (fats, sugars) present in the matrix. The Maillard reaction is the main chemical reaction that may affect the allergenicity of proteins by inducing the formation of aggregates which bind more effectively than unmodified allergens, and are also more resistant to gastric digestion.[41] Aggregated proteins are insoluble in non-denaturing conditions. The comparison of the spectra, unexpectedly revealed that the complete red PRVB thermal denaturation needs no less than 15 minutes of cooking (Figure 2-II). The increased thermostability of red PVRB may have been caused by conformational changes in the secondary structure.[42] Otherwise, body coloration is the unique discriminating factor between the analysed samples. Dietary supplementation of synthetic Xanthophylls are responsible for the red coloration in farmed trout which has the capability to transport and deposit these compounds at specific sites in the muscle. There have been speculations about where Xantophilles (Ax and Cx) bind in the muscle.[21,22,43-47] Further, Birkelanda et al. detected Ax in the water soluble fraction of salmon muscle (sarcoplasmatic fraction) indicating the formation of a water-soluble astaxanthin–protein complex.[48.] Accordingly, the observed increase in thermostability of parvalbumin from red rainbow trout might be related to an increased hydrophobicity associated to protein folding. Although the molecular mechanism related to the formation of protein-carotenoid complexes is not completely understood, the involvement of PRVB in pigment transport can be hypothesized.

PRVB variants from trout (Oncorhynchus mykiss).

PRVB possesses a rigid tertiary structure in metal-bound states, while the apo-form is intrinsically disordered, thus more accessible to the protease activity. Calcium depletion of the protein mixtures was performed by adopting analytical grade Chelex-100 ion-exchange resin in order to obtain a more extensive proteolytic digestion. MALDI MS spectrum of A_w, fraction from fresh fillet after calcium depletion displayed three iso-allergens with the molecular weights of 11280, 11772 and 11880 Da (Figure 3). Calcium induced changes in the protein conformation renders available amino acids that are buried in the apo-form of the protein. Thus, some of the proteins turn into differentially charged protein-species producing novel peaks (m/z) in MS experiments. Different PRVB variants have previously been reported in several fish species.[49,50]

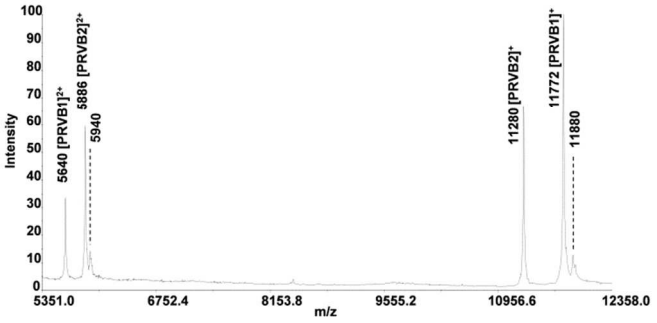


Figure 3. Linear MALDI MS spectrum of A_w, fraction from fresh fillet after calcium depletion.

Little information is available in the databases about the PRVBs from *Oncorhynchus mykiss*. Only two validated, incomplete isoforms for this species are included in the protein databases: Parvalbumin Beta 1 (P86431) and Parvalbumin Beta 2 (P86432). Sequence information was obtained by digesting fraction A and performing MS/MS on the digestion products, considering two different sets of experiments for white (A_w) and red (A_r) rainbow trout. The fraction A_w, was digested with the proteolytic enzymes: trypsin and proteinase K. The latter was chosen for its broad specificity since it cleaves polypeptides on the N-terminal side of aliphatic and aromatic amino acid residues in ammonium bicarbonate buffer (pH 8), generating peptides that are complementary to those produced by tryptic digestion. The direct submission of MALDI MS/MS data for protein identification was performed introducing

proteinase K for Enzyme on the Mascot Search Database Form (in house), and cleavage sites (P1: A, E, F, I, L, T, V and Y) as extra instructions in the Mascot "Enzyme file". MASCOT searches were performed against the SWISS PROT database, with the taxonomy restricted to Actinopterygii (ray-finned fishes), enzyme: proteinase K allowing 12 missed cleavages. Table 1 shows collectively peptide sequences produced by the proteinase K digestion which were identified.

Table 1. MS/MS identified peptides from PRVBs proteinase K digests.

	Sequence[a]	Mr found[b]	Mr calc[b]
Beta 1			
	KTALEA	632.37	632.36
	CKAADSF	743.38	741.32
	(Acetyl)ACAHLCKE	916.41	916.40
	TKAFLKAGDA	1021.57	1021.57
	KVIDQDASGF	1079.55	1079.54
	KAGDADGDGMIGIDEFAVL	1893.90	1893.89
	LKAGDADGDGMIGIDEFAVL	2006.99	2006.97
	LKAGDADGDGMIGIDEFAVL[c]	2028.97	2028.96
	FLKAGDADGDGmIGIDEFAVLVKQ	2525.28	2525.26
	GFASKSADDVKKAFKVIDQDASGF	2531.29	2531.28
	TDAETKAFLKAGDADGDGmIGIDEFA	2674.24	2674.22
Beta 2			
	AACTAA	507.23	507.22
	GLASKSSDDV[c]	1000.47	1000.46
	(Acetyl)SFAGLNDADVA[c]	1143.50	1143.49
	SARALTDAETKA	1233.65	1233.64
	AADSFNHKAFF	1254.60	1254.59
	YVIDQDKSGFI[c]	1306.64	1306.63
	LADGDKDGDGMIGV	1362.63	1362.62
	DGDKDGDGmIGVDE	1438.58	1438.56
	LQNFSASARALTDAETKA	1893.98	1893.97
	NHKAFFAKVGLASKSSDDV	2021.06	2021.04
	QNFSASARALTDAETKAFLA	2112.09	2112.07
	LADGDKDGDGMIGVDEFAAMI[c]	2161.96	2161.94

[a]Aminoacid sequence of peptides identified from proteinase K digests on the basis of their CID spectra. [b]All mass values are listed as monoisotopic mass. [c]Denotes $[M-Na]^+$ adduct. m Denotes methionine oxidized.

Table 2. MS/MS identified peptides from PRVBs trypsin digests.

	Sequence[a]	Mr found[b]	Mr calc[b]
Beta 1			
	TFFHTLGFASK	1255.66	1255.65
	VIDQDASGFIEVEELK	1791.91	1791.90
	VIDQDASGFIEVEELK[c]	1813.90	1813.88
	TFFHTIGFASKSADDVK[c]	1892.93	1892.92
	AGDADGDGMLGLDEFAVLVKQ	2121.03	2121.02
	AADSFNFKTFFHTLGFASK	2136.07	2136.06
	AGDADGDGMLGLDEFAVLVKQ[c]	2143.02	2143.00
	EADIKTALEACKAADTFSFK	2159.08	2159.07
	AFKVLDQDASGFLEVEELKc	2160.10	2160.08
	TFFHTLGFASKSSDDVKKAFK	2361.24	2361.22
	AFLKAGDADGDGMLGLDEFAVLVKQ	2580.32	2580.30
	AFLKAGDADGDGmLGLNEFAVLVKQ	2596.33	2596.31
	LFLQNFCPKARTLTDAETKAFLK	2655.45	2655.43
	AFLKAGDADGDGMLGLDEFAVWVKQ[c]	2675.30	2675.28
	VIDQDASGFIEVEELKLFLQNFCPK	2882.49	2882.46
	VIDQDASGFIEVEELKLFLQNFCPKAR	3109.63	3109.60
	TLTDAETKAFLKAGDADGDGMLGLDEFAVLVKQ	3439.76	3439.73
Beta 2			
	LFLQNFSAGAR	1223.66	1223.65
	TFFKACGLSPEEVK	1555.79	1555.78
	VFFLAKGAADLKVADAK	1764.02	1764.01
	AFYVIDQDKSGFIEEDELK	2246.10	2246.09
	SGFIEEDELKLFLQNFVAGAR	2383.25	2383.23
	AFLADGDKDGDGMIGVDEFAAMIKG	2543.20	2543.18
	LFLQNFSASARALSDAETKAFLK	2528.37	2528.35
	SAGSFDHTKFFKSCGLAGKSSDDVK	2619.27	2619.25
	ALTDAETKAFLADGDKDGDGMIGVDEFAAmIKG	3388.62	3388.59
	AFYVIDQDKSGFIEEDELKLFLQNFSASAR	3480.75	3480.73

[a]Aminoacid sequence of peptides identified from tryptic digests on the basis of their CID spectra. [b]All mass values are listed as monoisotopic mass. [c]Denotes $[M-Na]^+$ adduct. m Denotes methionine oxidized.

Peptides generated by trypsin-digested fraction ($A_{w'}$) were loaded on to the C18 column and a linear gradient elution was performed. All the collected fractions were vacuum concentrated and resuspended in 5 µL of CH_3CN/H_2O, 40:60 (v:v). A volume of 1 µL of the resulting solution was mixed with matrix and loaded on to the MALDI plate. Protonated tryptic peptides from each LC fraction were used to identify PRVB variants after MS/MS fragmentation. All tandem mass spectra were evaluated using MASCOT database searching. The oxidation of methionine, acetylation of protein N-term and sodium DE were included in the variable modifications. All results were carefully validated by manual interpretation of the corresponding MS/MS spectra. Table 2 lists the identified peptides for PRVBs.

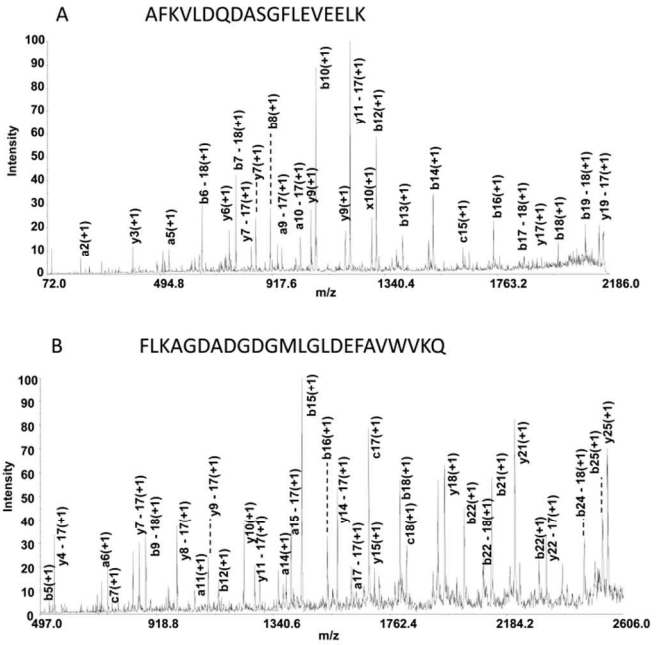


Figure 4. CID (1KeV) spectra of the sequences (A) AFKVLDQDASGFLEVEELK (m/z 2160.10), (B) AFLKAGDADGDGMLGLDEFAVWVKQ (m/z 2675.30) belonging to PRVB1 and variant PRVB1.1, respectively.

All amino acids were determined experimentally with the exception of leucine and isoleucine residues. These residues were assigned by comparing with the homologous proteins found in databases. Multiple adjacent cut sites can slow or block trypsin's activity. Therefore, peptide fragments with more than 2 missed cleavages were considered during mass analysis. CID spectra of peptides AFKVLDQDASGFLEVEELK (m/z 2160.10, A) and AFLKAGDADGDGMLGLDEFAVWVKQ (2675.30, B) are shown in Figure 4. These large peptides were valuable in assembling the protein sequences. For example, the large fragments of m/z 3109.63, identified by database search, and validated by MS/MS experiment as the sequence fragment VIDQDASGFIEVEELKLFLQNFCPKAR (Figure 5, T1) arose from two missed adjacent cut sites. It contains a large percentage of the total sequence and matchs two different unreviewed (TrEMBL) variants of PRVB_ONMCY (E0WDA2, E0WDA4). The same peptide contains the sequence fragments AFKVLDQDASGFLEVEELK (m/z 2160.10, Figure 5, T2), VIDQDASGFIEVEELK (m/z 1791.91, Figure 5, T3) and confirms the cleavage sites K(45), K(48) and R(75). The N-terminal residue AF in the peptide at m/z 2160.10 overlaps the AF residue found in peptide of m/z 2361.24 (TFFHTLGFASKSSDDVKKAFK, Figure 5, T4) at positions 46-47.

Journal Name RSCPublishing

ARTICLE

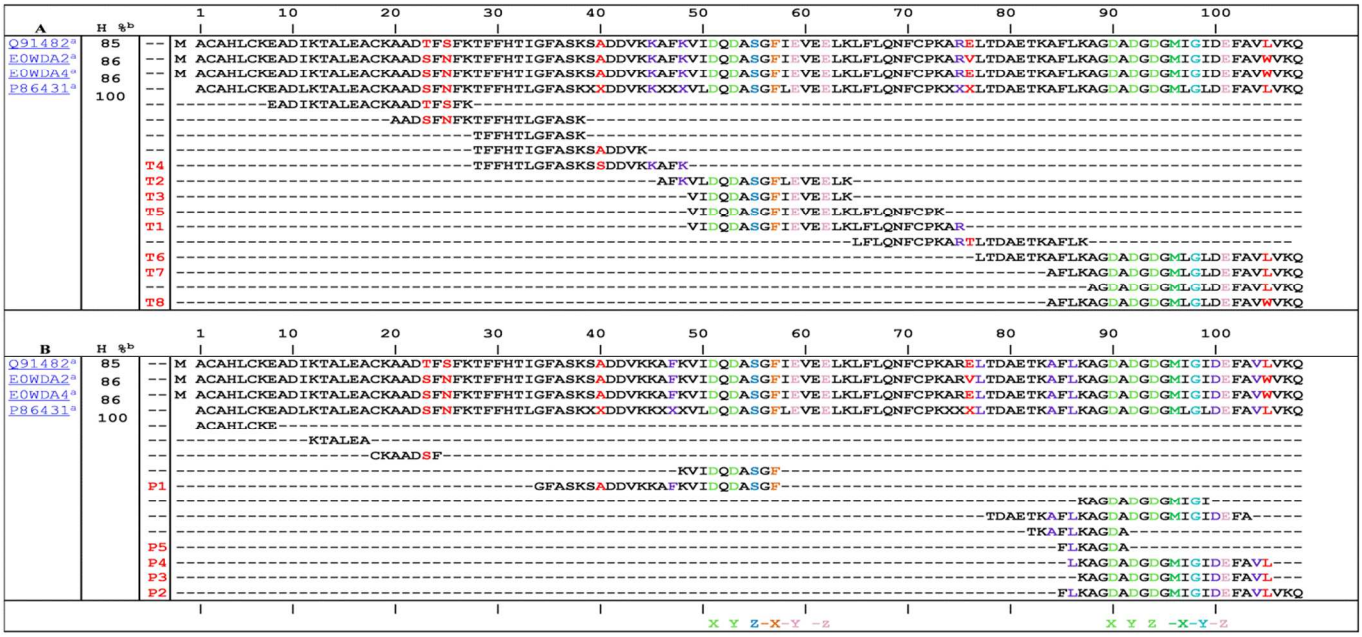


Figure 5. Alignment of the identified PRVB1 peptides from trypsin (panel A), proteinase K (panel B) digestion with sequences included in the protein databases. [a]According to "UniProtKB" (http://www.uniprot.org/). [b]Homology (%) according to "UniProtKB" http://www.uniprot.org/blast/uniprot/). X, Red and Blue denote unknown residues, point mutations and cleavage sites, respectively. Conservative CD domain (x, y, z , -x, -y –z) are highlighted using different colours

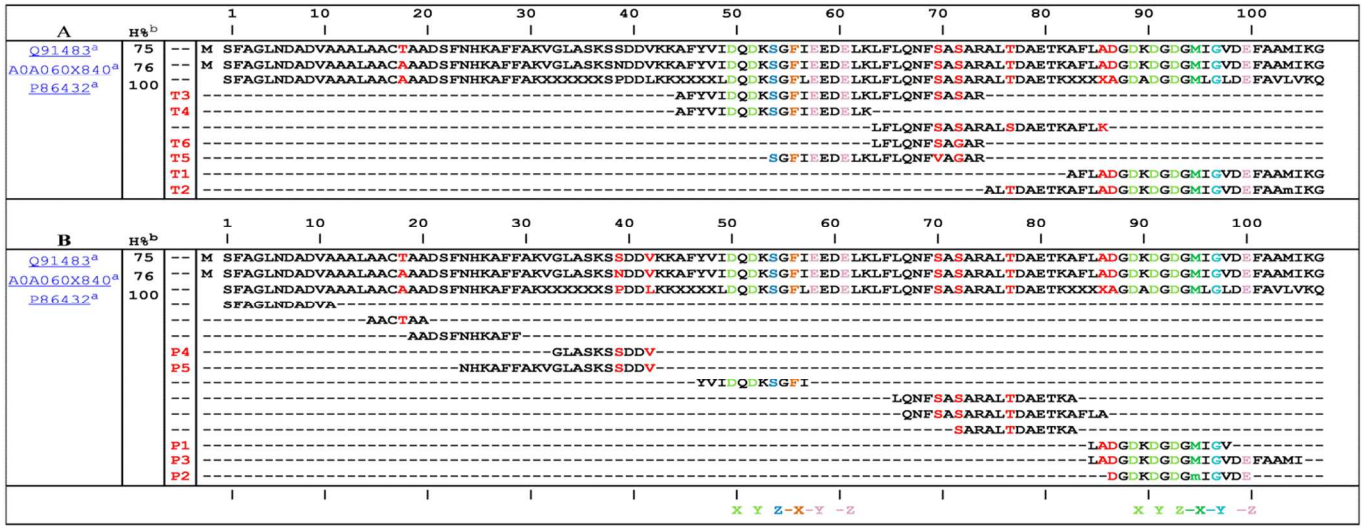


Figure 6. Alignment of the identified PRVB2 peptides from trypsin (panel A), proteinase K (panel B) digestion with sequences included in the protein databases. [a]According to "UniProtKB" (http://www.uniprot.org/). [b]Homology (%) according to "UniProtKB" http://www.uniprot.org/blast/uniprot/). X and Red denote unknown residues and point mutations, respectively. Conservative CD domain (x, y, z , -x, -y –z) are highlighted using different colours.

.

Thus, the amino acid sequence AF can be assigned at position 46-47 for all β1 components of the natural allergen mixture. Furthermore, the peptide of m/z 3109.63 (Figure 5, T1) covers the fragment sequence VIDQDASGFIEVEELKLFLQNFCPK (Figure 5, T5) plus C-terminal AR residue. The alignment of the two segments suggests that all β1 components of the mixture contain the residue AR at position 74-75. The sequence of m/z 2361.24 (Figure 5, T4) and m/z 3109.63 (Figure 5, T1), covering the middle region of the β1 variants, overlap the peptide GFASKSADDVKKAFKVIDQDASGF (m/z 2531.28,

Figure 5, P1) from proteinase K, thus the residue SA can be assigned at position 39-40 of all β1 variants.

Table 3. Deduced sequences, average calculated molecular masses and found pseudo-molecular ions of PRVBs.

	Sequence	MW[a]	$[M-H]^+$	$[M-2H]^{2+}$	$[M-Na]^+$	$[M-Na]^{2+}$
PRVB1	ACAHLCKEAD LKTALEACKA ADTFSFKTFF HTLGFASKSA DDVKKAFKVL DQDASGFLEV EELKLFLQNF CPKARTLTDA ETKAFLKAGD ADGDGMLGLD EFAVLVKQ	11771	11772	5886	--	--
PRVB1.1	ACAHLCKEAD LKTALEACKA ADSFNFKTFF HTLGFASKSA DDVKKAFKVL DQDASGFLEV EELKLFLQNF CPKARTLTDA ETKAFLKAGD ADGDGMLGLD EFAVWVKQ	11857	--	--	11880	5940
PRVB2	SFAGLNDADV AAALAACTAA DSFNHKAFFA KVGLASKSSD DVKKAFYVID QDKSGFLEED ELKLFLQNFS ASARALSDAE TKAFLKDGDA DGDGMLGVDE FAAMIKG	11280	11281	5640	--	--
PRVB2.1	SFAGLNDADV AAALAACTAA DSFNHKAFFA KVGLASKSSD DVKKAFYVID QDKSGFLEED ELKLFLQNFS ASARALTDAE TKAFLADGDA DGDGMLGVDE FAAMIKG	11236	--	--	--	--

[a]Denotes average calculated masses.

The exchange of one amino acid at the position 76 (E→T) leads to a mass difference of -27.99 Da and allows to attribute the sequence TLTDAETKAFLKAGDADGDGMLGLDEFAVLVKQ (Figure 5, T6) to the experimental mass 3439.76 Da with 9 ppm error. This sequence is supported by the fragmentation pattern of the corresponding $[M-H]^+$ species matching all peaks visible in the spectrum. Furthermore, a mass difference of 95 Da between the peaks at m/z 2580.32 (AFLKAGDADGDGMLGLDEFAVLVKQ, Figure 5, T7) and m/z 2675.30 (AFLKAGDADGDGMLGLDEFAVWVKQ, Figure 5, T8) can be justified by the exchange of one amino acid at position 105 (L→W), indicating the presence of a new variant (PVRB1.1). The MS/MS data confirm the proposed sequence AFLKAGDADGDGMLGLDEFAVWVKQ (Figure 5, T8), matching the segment 84-108 in E0WDA2 and E0WDA4 variants.

The peptide AFLKAGDADGDGMLGLDEFAVLVKQ (m/z 2580.32, Figure 5, T7) provides overlap between some proteinase K peptides (Figure 5, P2-P5). In particular, peptide of m/z 2525.28 (Figure 5, P2) has identical sequence differing only for an F residue at the N-terminus of the peptide (incomplete digestion product). Thus the C-terminal region of PVRB1 is made up of these sequence fragments. Peptides resulting from digestion with trypsin end in either lysine or arginine, whereas peptides originating from the C-terminus of the proteins may end in any amino acid. Construction of the PRVB2 sequence started from the protonated tryptic peptides that do not end in lysine or arginine. The peptides of m/z 2543.20 (AFLADGDKDGDGMIGVDEFAAMIKG, Figure 6, T1) and m/z 3388.62 (ALTDAETKAFLADGDKDGDGMIGVDEFAAmIKG, Figure 6, T2) were assigned to the C-termini of PRVB2. The last large peptide overlaps the peptides of m/z 1362.63, 1438.58 and 2161.96 (Figure 6, P1-P3) from proteinase K, thus the residue AFLA can be assigned at position 83-86 of all β2 variants. The alignment of the fragment peptides found at m/z 3480.75, 2246.10, 2383.25 and 1223.66 (Figure 6, T3-T6) allows to cover the middle region of PRVB2, assigning the residues AFYVI at position 45-49. Furthermore, the peptides found at m/z 2383.25 (SGFIEEDELKLFLQNFVAGAR, Figure 6, T5) and 1223.66 (LFLQNFSAGAR, Figure 6, T6) suggest the presence of additional point mutations S→V and S→G at position 70 and 72, respectively. Finally, the observed peptides at m/z 1000.47 (GLASKSSDDV, Figure 6, P4) and 2021.0451 (NHKAFFAKVGLASKSSDDV, Figure 6, P5) cover the last gap of PRVB2 sequence and assign the residue VGLASK at position 32-37. The primary structure of PRVBs is conserved over a wide phylogenetic range demonstrating a high content of conservative regions and severe sequence similarities. Therefore, all peptides identified were aligned with sequences included in the protein databases showing the highest degree of homology (up to 75%). The alignments for PRVBs are shown in Figure 5 and 6. The total protein sequences, average calculated masses and average measured masses for PRVB1, PRVB1.1, PRVB2 and PRVB2.1 are shown in Table 3. The N-terminal residue was found to be acetylated in all cases. As shown in Table 3, an excellent agreement was found between theoretical and experimental masses for the PRVBs yielding accuracies in the Mr measurements. Taking into account the common amino acid positions identified, the analysis of PRVB1 sequence shows three point mutations confirming the presence of two isoforms (PRVB1 and PRVB1.1). Therefore, the found peaks at m/z 11772 and 5886 are $[PRVB1-H]^+$ and $[PRVB1-H]^{2+}$, respectively (Figure 3). The exchange of the amino acids at the position (23) T→S, (25) S→N and (105) L→W leads to a difference of 85.99 Da, then the ion peaks of m/z 11880 and 5940 correspond to $[PRVB1.1-Na]^+$ and $[PRVB1.1-Na]^{2+}$, respectively (Table 3, Figure 3). The ions found at m/z 11280 and 5640 correspond to $[PRVB2-H]^+$ and $[PRVB2-H]^{2+}$, respectively. The observed point mutations at (77) S→T and (86) K→A indicate the presence of a new low abundant variant in the natural mixture PRVB2.1 (Table 3). No additional point mutations were observed for PRVBs from red rainbow trout.

PRVB belongs to a family of proteins with several highly conserved helix-loop-helix (EF-hand) motifs[51] which bind both Ca^{2+} and Mg^{2+}. EF-hand domain consists of a twelve residue loop flanked on both sides by a twelve residue α-helical domain.[52] In an EF-hand loop the calcium ion is usually coordinated in a pentagonal bipyramidal configuration. The six residues involved into the metal coordination are in positions 1, 3, 5, 7, 9 and 12, denoted by X, Y, Z, -X, -Y and -Z, respectively (Figures 5, 6). Positions 1 (X), 3 (Y) and 12 (-Z) are the most conserved, the last one being usually glutamic acid (E) or aspartic acid (D). In most cases the 6th residue is glycine (G). The signature pattern includes the complete EF-hand loop as well as the first hydrophobic residue which follows the loop. In reconstructed PRVB sequences, there are leucine (L) and phenylalanine (F) that follow EF-hand1 and EF-hand2, respectively (Figure 5, 6). In the CD domain, all residues localized in the Ca^{2+}-binding positions are conserved (X: Asp, Y: Asp, Z: Ser, -X: Phe, -Y: Glu, -Z: Glu). The same for the EF domain, (X: Asp, Y: Asp, Z: Asp, -Y: Gly, -Z: Glu), except for the position (-X), in which three possible different residues (Lys, Ala, Met) can be found. The reconstructed PRVB sequences show a methionine residue at -X position as deduced by the identified sequences containing the oxidized residue.

Conclusions

A top-down based strategy was successfully used for complete sequencing of PRVBs variants from rainbow trout. The results showed that the combined use of selective solubilization, specific digestion and MALDI MS/MS measurements provide an unambiguous determination of the sequence, point mutation of the examined low-molecular weight and thermostable proteins. An important role was played by the use of proteinase K digestion-based identification because the enzyme cleaves polypeptides on the N-terminal side of aliphatic and aromatic aa that are complementary to those produced by tryptic digestion. Special attention was taken while to preparing protein extracts from cooked flesh muscle, simulating the way consumers are usually exposed to food allergens. Molecular profiling showed that thermal processing of red flesh muscle of rainbow trout could not limit allergic reactions in humans.

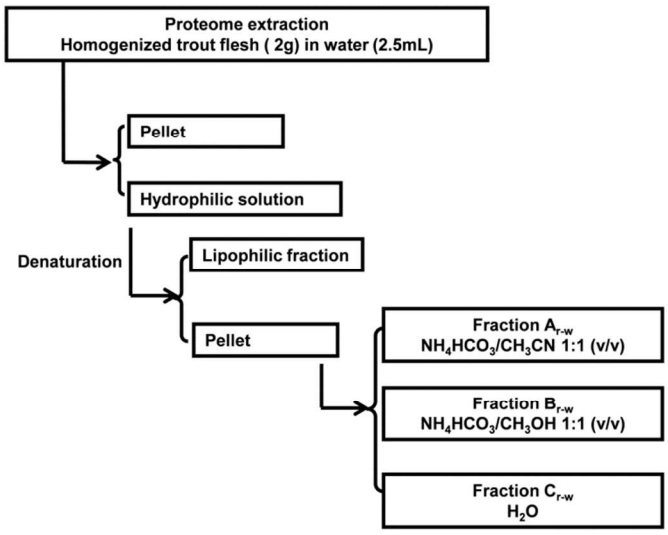
Experimental

Materials.

White and red rainbow trout (*Oncorhynchus mykiss*) were purchased from a local trout farm. Solvents (CH_3OH, CH_3CN, $CHCl_3$, and H_2O, HPLC grade), ammonium bicarbonate (NH_4HCO_3, pure 99.5%), trifluoroacetic acid (pure 99.0%), enzymes (trypsin, proteinase K), and α-cyano-4-hydroxy-trans-cynnamic acid (α-CHCA, pure 99.0%) were purchased from Sigma Aldrich Fluka (Milano, Italy). Chelex chelating ion exchange resin (Chelex 100 Resin, biothecnology grade 100-200 mesh, sodium form) was purchased from BIO-RAD (Hercules, CA).

Extraction of sarcoplasmic proteins.

Fresh fillets from white and red rainbow trout were homogenised with a blender, than stored at -20°C. Sarcoplasmic proteins extraction was carried out using 2g of homogenised white or red trout flesh with 2.5 mL of water, under magnetic stirring for 20 min at room temperature followed by a centrifugation at 6000 rpm for 2 min. The supernatant portion was collected and clarified by centrifugation at 12000 rpm for 1 min. A 200 µL portion of whole extract was treated with 1 mL of $CHCl_3/CH_3OH$ 1:3 (v/v), and the obtained pellet was partitioned consecutively under magnetic stirring at room temperature as reported.[53,54]



Scheme. schematic workflow

To test PRVB thermal stability, 2 g portions of homogenized white and/or red trout fleshes were cooked by bain-marie mode at 100°C, for 3, 5, 10 and 15 min. Then cooked flesh samples were handled as previously described.[53,54] Each fraction was directly analysed by linear MALDI MS.

PRVB calcium depletion.

A 400µL portion of fractions A_r and A_w (Scheme) for white and red trout respectively, was dried in a speed-vacuum concentrator system (Concentrator Plus - Eppendorf, Hamburg-Germany) and then solubilised with 500 µL of Milli-Q filtered deionized water. To remove calcium ions from PRVB, an analytical grade Chelex-100 ion-exchanger was used applying batch technique. A 50 mg portion of gel-type resin was equilibrated in water and then added to $A_{r'}$ and $A_{w'}$ fractions. The ion exchange was performed under gently shaking overnight, and the free calcium samples ($A_{r'}$ and $A_{w'}$) were collected.

PRVB proteolytic digestion.

A 50 µL portion of free calcium samples $A_{r'}$ and $A_{w'}$ were completely dried in speed vacuum, the pellets were solubilised in 50 µL of 50 mM NH_4HCO_3 (pH~8) and added with 5 µL (0.4 pmol/µL) of trypsin. The completed microwaves assisted digestion was obtained after five simultaneous treatments in the microwave (MWD 246 SL, Whirlpool Europe, Italy) at 250 W irradiation power each lasting for 3 min. Separation of the peptides derived from the $A_{r'}$ and $A_{w'}$ digests were carried out using HPLC (Agilent 1100 Germania) equipped with a C18 column (Hypersyl Gold 100 mm×4.6 mm i.d., Thermo, Bellefonte, PA). Dried peptides were dissolved in 50 µL of TFA 0.1% and injected. The flow rate was set at 1 mL/min by using the following gradients: solvent A (H_2O, 0.1% TFA), solvent B (CH_3CN) from 10% B to 100% B in 35 min, 2 min at 100% B isocratic.[55] The column effluent was collected at 0.5 min intervals resulting in 80 fractions. Each of the RP-LC fractions were completely dried by Speed vac and resuspended in 5 µL of CH_3CN/H_2O, 40:60 (v:v).

A 50 µL portion of free calcium sample ($A_{w'}$ and $A_{r'}$) was digested by proteinase K as described,[56] and a 1 µL portion of digests was mixed with 10 µL of α-CHCA.

Protein identification by MALDI MS and MS/MS analysis.

Measurements of molecular mass (Mr) of intact PRVBs were performed on a 4700 Proteomics Analyzer mass spectrometer from Applied Biosystems (Foster City, CA) equipped with a 200-Hz Nd:YAG laser at 355-nm wavelength. Linear MALDI MS spectra were acquired averaging 3000 laser shots with a mass accuracy of 500 ppm in default calibration mode that was performed using a set of standard proteins (insulin bovine, $[M+H]^+$ avg. 5734.59 apomyoglobin horse, $[M+H]^{2+}$ avg. 8476.78, $[M+H]^+$ avg. 16952.56, thioredoxin E. coli, $[M+H]^+$ avg. 11674.48). A 1-µL portion of each fraction was directly analyzed by linear MALDI using α-cyano-4-hydroxy-transcynnamic acid (α-CHCA, 0.3% in TFA) as matrix.

A 1 µL portion of protein digests matrix solution was spotted on a MALDI matrix target, dried at room temperature, and directly analyzed by 5800 MALDI TOF/TOF analyzer (AB SCIEX, Darmstadt, Germany) equipped with a neodymium: yttrium-aluminum-garnet laser (349 nm), in reflectron positive mode with a mass accuracy of 5 ppm. At least 4000 laser shots were typically accumulated with a laser pulse rate of 400 Hz in the MS mode, whereas in the MS/MS mode spectra up to 5000 laser shots were acquired and averaged with a pulse rate of 1000 Hz. MS/MS experiments were performed at a collision

energy of 1 kV, and ambient air was used as the collision gas with a medium pressure of 10^{-6} Torr. After acquisition, spectra were handled using Data Explorer version 4.11 (AB Sciex). The MS/MS data were also processed to assign candidate peptides in the SwissProt database using the MASCOT search program (http://www.matrixscience.com). All MS/MS searches were performed using an initial mass tolerance of 50 ppm and the query was made for "Actinopterygii (ray-finned fishes)" Taxonomy, allowing at most three missed cleavages for tryptic peptides. The mass tolerance of the fragments for MS/MS data search was set at 0.20 Da (Figure. MS/MS spectra that were not identified by MASCOT were analyzed by de novo peptide sequencing using Data Explorer. All peptide sequences were later corroborated by BLAST (http://www.uniprot.org/blast/), showing in all cases homologies with other PRVBs present in the databases. All peptide sequences were arranged by overlapping the results obtained with both enzymatic digests and by comparison with the proteins included in the UniProtKB database. The theoretical Mr was calculated using the Compute pI/Mw tool (http://web.expasy.org/compute_pi/).

Electronic Supporting Information Available:

Linear MALDI MS spectra of fractions Bw, Ar, Br, Cr (Figure 1-3ESI). MS/MS based sequence identification by MASCOT search of the ion species of m/z 2121, 2580, 1255 and 1791 (Figure 4-7ESI).

Acknowledgements

We thank the European Union for the financial support through the project CHEBANA (FP7-PEOPLE-2010-ITN-264772).

Notes and references

[a]Department of Chemistry and Chemical Technologies, University of Calabria, Italy
[b]Department of Chemistry, "Sapienza" University of Rome, Rome, Italy
[c]Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Italy
*Corresponding Author: Prof. Anna Napoli, Department of Chemistry and Chemical Technologies, University of Calabria, Via P. Bucci, Cubo 12/D, 87036 Arcavacata di Rende (CS), Italia. e-mail: amc.napoli@unical.it; Fax: +39 0984 493307

1 S. H. Sicherer, *J. Allergy Clin. Immunol.* 2011, **27**, 594−602.
2 C.Y Pascual, M. Reche, A. Fiandor ,T. Valbuena, T. Cuevas, M. M. Esteban, *Allergy Immunol.* 2008, **19**, 573−579.
3 P. Failler, Future prospects for fish and fishery products.4. Fish consumption in the European Union in 2015 and 2030 FAO Fisheries Circular No. 972/4 Food and Agriculture Organization of the United Nations Rome, 2007.
4 M. F. Sharp, A. L. Lopata, *Clin. Rev. Allergy Immunol.* 2014, **46**, 258–271.
5 M. F. Jeebhay, T. G. Robins, S. B. Lehrer, A. L. Lopata, *Occup. Environ. Med.* 2001, **58**, 553–562.
6 H. A. Sampson, *Pediatrics* 2003, **111**, 1601–1608.
7 C. D. Lindstrøm, T. van Dô, I. Hordvik, C. Endresen, S. Elsayed, *Scand. J. Immunol.* 1996, **44**, 335–344.
8 J. P. Declercq, B. Tinant, J. Parello, *J. Mol. Biol.* 1991, **220**, 1017–1039.
9 J. P. Declercq, C. Evrard, V. Lamzin, J. Parello, *Protein Sci.* 1999, **8**, 2194–2204.
10 Y. Ma, U. Griesmeier, M. Susani, C. Radauer, P. Briza, A. Erler, M. Bublin, S. Alessandri, M. Himly, S. Vàzquez-Cortés, I. R. de Arellano, E. Vassilopoulou, P. Saxoni-Papageorgiou, A. C. Knulst, M. Férnndez-Rivas, K. Hoffmann-Sommergruber, H. Breiteneder, *Mol. Nutr. Food Res.* 2008, **52**, S196–S207.
11 M. Koeberl, D. Clarke, A. L. Lopata, *J. Proteome Res.* 2014, **13**, 3499–3509.
12 M. Carrera, B. Cañas, C. Piñeiro, J. Vázquez, J. M. Gallardo, *Proteomics* 2006, **6**, 5278–5287.
13 J. E. Beale, M. F. Jeebhay, A. L. Lopata, *Mol. Immunol.* 2009, **46**, 2985–2993.
14 A. Kobayashi, H. Tanaka, Y. Hamada, S. Ishizaki, Y. Nagashima, K. Shiomi, *Allergy* 2006, **61**, 357–363.
15 A. Bugajska-Schretter, E. L. Fuchs, T. S. Kapiotis, H. Rumpold, R. Valenta, S. Spitzauer, *J. Allergy Clin. Immunol.* 1998, **101**, 67–74.
16 Scientific Opinion of the Panel on Animal Health and Animal Welfare on a request from the European Commission on the Animal welfare aspects of husbandry systems for farmed trout. *The EFSA Journal* 2008, **796**, 1-22.
17 Regulation (EC) No 1831/2003 of the European Parliament and the Council of 22 September 2003 on additives for use in animal nutrition. Official Journal of the European Union, L268/29, 18.10.2003
18 Commission regulation (EC) No 880/2004 of 29 April 2004 authorising without time limit the use of beta-carotene and canthaxanthin as additive in feedingstuffs belonging to the group of colouring matters including pigments. Official Journal of the European Union, L162/68, 30.4.2004.
19 Commission regulation (EC) No 1288/2004 of 14 July 2004 concerning the permanent authorisation of certain additives and the D.E. Breithaupt / Trends in Food Science & Technology 18 (2007) 501e506)
20 T. Ytrestøyl, G. Struksnaes, W. Koppe, B. Bjerkeng, *Comp. Biochem. Physiol. Part B:Biochem. Mol. Biol.* 2005, **142**, 445–455.
21 T. Ytrestøyl, G. Coral-Hinostroza, B. Hatlen, D. H. Robb, B. Bjerkeng, *Comp. Biochem. Physiol. Part B:Biochem. Mol. Biol.* 2004, **138**, 29–40.
22 H. Henmi, M. Hata, M. Hata, *Nippon Suisan Gakkaishi*. 1990, **56**, 1821-1823.
23 G. Britton, *FASEB J.* 1995, **9**, 1551–1558.
24 F. Mazzotti, L. Di Donna, D. Taverna, M. Nardi, D. Aiello, A. Napoli, G. Sindona, *Int. J. Mass Spectrom.* 2013, **352**, 87−91.
25 L. Di Donna, H. Benabdelkamel, F. Mazzotti, A. Napoli, M. Nardi, G. Sindona, *Anal. Chem.* 2011, **83**, 1990−1995.
26 F. Mazzotti, L. Di Donna, A. Napoli, D. Aiello, C. Siciliano, C. M. Athanassopoulos, G. Sindona, *J. Mass Spectrom.* 2014, **49**, 802-810.
27 F, Mazzotti, H. Benabdelkamel, L. Di Donna, L. Maiuolo, A. Napoli, G. Sindona, *Food Chem.* 2012, **135**, 1006−1010.
28 F. Mazzotti, H. Benabdelkamel, L. Di Donna, C. M. Athanassopoulos, A. Napoli, G. Sindona, *J. Mass Spectrom.* 2012, **47**, 932−939.
29 A. Napoli, C. M. Athanassopoulos, P. Moschidis, D. Aiello, L. Di Donna, F. Mazzotti, G. Sindona, *Anal. Chem.* 2010, **82**, 5552−5560.
30 V. Cunsolo, R. Saletti, V. Muccilli, S. Foti, *J. Mass Spectrom.* 2007, **42**, 1162-1174.
31 A. Napoli, D. Aiello, L. Di Donna, P. Moschidis, G. Sindona, *J. Proteome Res.* 2008, **7**, 2723−2732.
32 A. Napoli, D. Aiello, G. Aiello, M. S. Cappello, L. Di Donna, F. Mazzotti, S. Materazzi, M. Fiorillo, G. Sindona, *J. Proteome Res.* 2014, **13**, 2856–2866.
33 N.L. Kellehel, *Anal. Chem.* 2004, **76**, 197A–203A.

34 J.R. Whiteaker, B. Warscheid, P. Pribil, Y. Hathout, C. Fenselau, *J. Mass Spectrom.* 2004, **39**, 1113–1121.
35 S. H. Arif, M. Jabeen, A.-U. Hasnain, *J. Food Biochem.* 2007, **31**, 121–137.
36 M. B. Carrera, J. Cañas, M. Gallardo, *J. Proteomics* 2013, **78**, 211–220.
37 M. B. Carrera, J. Vázquez, J. M. Gallardo, *J. Proteome Res.* 2010, **9**, 4393–4406.
38 L. C. Delbarre, V. Verrez-Bagnis, J. Noel, J. Fleurence, *Food Chem.* 2003, **81**, 517–525.
39 A. Bugajska-Schretter, M. Grote, L. Vangelista, P. Valent, W. R. Sperr, H. Rumpold, A. Pastore, R. Reichelt, R. Valenta, S. Spitzauer, *Gut* 2000, **46**, 661–669.
40 K. Thomas, C. Herouet-Guicheney, G. Ladics, G. Bannon, A. Cockburn, R. Crevel, J. Fitzpatrick, C. Mills, L. Privalle, S. Vieths, *Food Chem. Toxicol.* 2007, **45**, 1116–1122.
41 S. D. Kamath, A. M. Abdel Rahman, T. Komoda, A. L. Lopata, *Food Chem.* 2013, **141**, 4031–4039
42 M. Albrecht, Y. Kuhne, B. K. Ballmer-Weber, W. M. Becker,. T. Holzhauser, I. Lauer, *J. Allergy Clin. Immunol.* 2009, **124**, 328–336.
43 H. Henmi, T. Iwata, M. Hata, M. Hata, *Tohoku J. Agric. Res.* 1987, **37**, 101–111.
44 S. J. Matthews, N. W. Ross, S. P. Lall, T. A. Gill, *Comp. Biochem. Physiol. Part B* 2006, **44**, 206–214.
45 H. Henmi, M. Hata, & M. Takeuchi, *Comparative Biochemistry and Physiology*, 1991, **99B**, 609–612.
46 G. H. Aas, B. Bjerkeng, T. Storebakken, & B. Ruyter, *Fish Physiology & Biochemistry*, 1999, **21**, 325–334.
47 L. Byrnes, & F. Gannon,. *DNA & Cell Biology*, 1990, **9**, 647–655.
48 S. Birkelanda, B. Bjerkeng, Food Chem. 2004, **85**, 559–568
49 K. G. Gajewski, Y. H. Hsieh, *J. Food Prot.* 2009, **72**, 818−825.
50 T. van Do, I. Hordvik, C. Endresen, S. Elsayed, *Mol. Immunol.* 2005, **42**, 345−353.
51 M. Ikura, *Trends Biochem. Sci.* 1996, **21**, 14–17.
52 H. Kawasaki, R. H. Kretsinger *Protein Profile* 1994, **1**, 343–517.
53 A. Napoli, D. Aiello, L. Di Donna, A. Sajjad, E. Perri, G. Sindona, *Anal. Chem*. 2006, **78**, 3434−3443.
54 A. Napoli, D. Aiello, L. Di Donna, H. Prendushi, G. Sindona, *Anal. Chem.* 2007, **79**, 5941−5948.
55 C. L. Di Donna, F. Mazzotti, D. Taverna, A. Napoli, G. Sindona, *Phytochem. Anal.* 2014, **25**, 207−212.
56 E. Reddy, L. Albanito, P. De Marco, D. Aiello, A. Napoli, A. M. Musti, *Cell Death Dis.* 2013, **4**, e852; doi:10.1038/cddis.2013.381