



Sequence microheterogeneity of parvalbumin, the major fish allergen



Yulia S. Lapteva^a, Vladimir N. Uversky^{a,b,*}, Sergei E. Permyakov^{a,c,**}

^a Protein Research Group, Institute for Biological Instrumentation of the Russian Academy of Sciences, Pushchino, Moscow region 142290, Russia

^b Department of Molecular Medicine and Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

^c Department of Biomedical Engineering, Pushchino State Institute of Natural Sciences, Pushchino, Moscow region 142290, Russia

ARTICLE INFO

Article history:

Received 18 February 2013

Received in revised form 5 April 2013

Accepted 21 April 2013

Available online 28 April 2013

Keywords:

Structural microheterogeneity

Protein polymorphism

Protein isoforms

Allergen

EF-hand

Parvalbumins

Oncomodulin

Intrinsic disorder

ABSTRACT

The microheterogeneity of amino acid sequence observed in various allergens may affect immune response, but incidence of sequence microheterogeneity in allergens and its relation to their allergenicity are unclear. The occurrence of sequence microheterogeneity in major fish allergen, parvalbumin (PA), has been explored using bioinformatics approaches. 44% of 111 species with known PA sequence have PA isoforms. 41% of these species exhibit from 1 to 4 cases of PA sequence microheterogeneity, i.e. unique pairs of PA isoforms with sequence identity above 90%. 29% of 210 PA sequences studied are characterized by microheterogeneity. The occurrence of allergens among them is 2.5-fold higher than among other PAs. The incidence of sequence microheterogeneity within more allergenic β isoform of PA is 2.0-fold lower than that for its less allergenic α isoform. 39 residues affected by PA microheterogeneity are concentrated in the region of helices A, B, F, while helices D and E are the most conservative region. 44% and 11% of the microheterogeneous substitutions are located in the species-specific and cross-reactive IgE-binding epitopes of PAs, respectively. 45% and 48% of the substitution cases are predicted to cause notable changes in protein disorder propensity and protein stability, respectively. Hence, the increased allergenicity rate among PAs having microheterogeneous isoforms can be related to differences in their IgE-binding caused directly or in an allosteric manner. Overall, sequence microheterogeneity is shown to be inherent to many of PAs and likely contributes to PA allergenicity.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Minor differences in protein's amino acid sequence (referred to as “sequence microheterogeneity”) may arise either due to evolutionary processes (gene duplication and divergence, appearance of allelic variants of a gene), alternative splicing or RNA editing. Since these mechanisms ensure functional diversification of proteome, sequence microheterogeneity is inherent to proteins, including allergenic ones. For example, multiple protein isoforms with pairwise sequence identity above 90% (will refer to as “microheterogeneous isoforms”) are found in major cow's milk allergen Bos d 5 (WHO/IUIS Allergen Nomenclature [1]), major apple allergen Mal d 1, major olive pollen allergen Ole e 1, major hazel pollen allergen Cor a 1, major Timothy grass pollen allergens Phl p 5 and Phl p 6, major birch pollen allergen Bet v 1 [2], and major house dust mite allergens Der p 1, Der p 2, Der f

1 and Der f 2 [3]. The cutoff of 90% is chosen since polymorphic variants of allergens show greater than 90% sequence identity [1]. Importantly, the term “sequence microheterogeneity” is used here without regard to the origins of this phenomenon.

The sequence microheterogeneity observed in various allergens reflects evolution of these proteins, but may represent a structural feature important for their allergenicity. The latter suggestion is supported by comparative studies of microheterogeneous isoforms of several allergens. Human IgE and T cell reactivity patterns, as well as IgE and IgG responses of microheterogeneous isoforms of Bet v 1 may greatly differ [4,5]. Similarly, prominent differences in IgE binding and potencies in inducing T cell activation were reported for recombinant microheterogeneous isoforms of Cor a 1 (see [6,7]) and Der p 2 [8,9]. Similar effects were observed for Ole e 1-like family [10]. Even single amino acid substitution may alter immune response, as evidenced by different cytokine patterns of T cell responses induced by recombinant Der p 2 isoforms 2.0107 and 2.0108 (V → A substitution) [8]. Similarly, a K → E substitution in dust mite allergen Lep d 2.02 resulted in strongly reduced IgE binding [11].

In contrast to microheterogeneous isoforms of allergens, sequence identities of isoallergens may reach 67% [1], while homologous allergens from the same species may have much lower sequence identity. Hence, comparable immune responses can be elicited by wide repertoire of molecular forms of the same allergen. Furthermore, allergens from different sources may be cross-reactive, which in common cases

Abbreviations: PA, parvalbumin; α -PA, α isoform of parvalbumin; β -PA, β isoform of parvalbumin

* Correspondence to: V.N. Uversky, Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL 33612, USA. Tel.: +1 813 974 5816; fax: +1 813 974 7357.

** Correspondence to: S.E. Permyakov, Institute for Biological Instrumentation of the Russian Academy of Sciences, Institutskaya str., 7, Pushchino, Moscow region 142290, Russia. Tel.: +7 4967 31 89 72; fax: +7 4967 33 05 22.

E-mail addresses: vuversky@health.usf.edu (V.N. Uversky), permyakov.s@gmail.com (S.E. Permyakov).

have been explained by structural similarities between the allergens [12]. Taking into account the wide range of possible structural differences between isoforms of an allergen and its cross-reactive relatives from other sources, sequence microheterogeneity may be considered as generally insignificant accompanying property of some allergens, emerging due to evolutionary processes. Meanwhile, the ability of minor changes in allergen's amino acid sequence to affect immune response leaves a chance that presence of microheterogeneous isoforms of an allergen may be important for its allergenicity. Unfortunately, the abundance of sequence microheterogeneity among allergens has not been properly explored, excluding some evident cases of complete predominance of microheterogeneous isoforms (for example, cases of Bos d 5, and Der p 1, Der p 2, Der f 1, see [3]). In the present work the occurrence of sequence microheterogeneity in major fish allergen, parvalbumin, is scrutinized.

Parvalbumin (PA) is a vertebrate-specific, mainly cytosolic, small (106–113 residues) globular acidic (pI 3.9–5.5) Ca^{2+} -binding protein of the EF-hand superfamily, which is expressed in fast-twitch muscles, specific neurons, certain kidney and endocrine glands cells, Corti's cells, and others [13]. In fast muscles, PA serves as a soluble relaxing factor accelerating Ca^{2+} -mediated muscle relaxation (for review, see [14]), while in early distal tubule PA is engaged into NaCl reabsorption [15]. Two phylogenetic lineages of PA, α and β , differ in isoelectric point (α : pI > 5; β : pI < 5) and multiple peculiarities of amino acid sequence [16], but share similar tertiary structure: three 30-residue-long subdomains, each containing a central loop flanked by short amphipathic α -helices (labeled A–F) [13,17]. Helices C, D and E, F with their interconnecting loops form two Ca^{2+} / Mg^{2+} -binding motifs of the EF-hand type. The shortened AB-loop prevents AB subdomain from effective metal binding.

Parvalbumin is a major cross-reactive fish allergen [18–23]. It represents the second largest animal food allergen family [24]. According to the WHO/IUIS Allergen Nomenclature Sub-committee, 18 PAs are allergens [1]. In individuals sensitized to PA a consumption or contact with fish-related products may cause urticaria and dermatitis, angioedema, diarrhea, asthma, systemic anaphylactic reactions and death (for review, see [20,23]). The high structural stability of Ca^{2+} -loaded PAs seems to promote allergic reactions despite fish cooking and exposure to the gastrointestinal tract [25]. High PA expression levels also favor its allergenicity [21,26].

Fish species have been shown to display up to eleven (African catfish [27]) PA isoforms, presumably favoring physiological adaptation of the propulsive musculature of fish to developmental and/or environmental requirements [27,28]. Our previous study of northern pike α -PA has shown that it exhibits microheterogeneity of amino acid sequence [29]. The analysis of available PA sequences presented here evidences that sequence microheterogeneity is fairly common for parvalbumins. The presence of sequence microheterogeneity is shown to correlate positively with PA's allergenicity. The analysis of amino acid substitutions accompanying sequence microheterogeneity in PA enables to suggest a structural rationale for this observation. Since sequence microheterogeneity might be a general structural feature promoting allergenicity of food allergens, an alternative explanation of higher incidence of allergic diseases among population of developed countries and urban areas is proposed. Evolutionary aspects of sequence microheterogeneity in PA are also discussed.

Though calcium-saturated PAs are well-folded and highly stable proteins, recent study showed that the apo-forms of pike α - and β -PAs behaved as compact intrinsically disordered proteins that substantially devoid fixed tertiary structure under the physiological conditions *in vitro*, suggesting that calcium removal was sufficient to transform these proteins into the molten globule-like conformations [30]. Therefore, it is likely that the predisposition of parvalbumins for intrinsic disorder might contribute to their allergenicity. These observations are in line with the recent finding that many of the allergen representative peptides (ARPs) are expected to be disordered [31]. In fact,

bioinformatics analysis revealed that about 20% of the allergen peptides were predicted to be highly disordered, and that about 77% of ARPs were either located within disordered regions of corresponding allergenic proteins or showed more disorder/flexibility than their neighbor regions [31]. Furthermore, in a subset of the allergenic proteins, a sequence co-localization between the ARPs and the molecular recognition features (MoRFs, which are short interactive disordered regions undergoing disorder-to-order transitions at interaction with binding partners) was found, suggesting that intrinsic disorder and MoRFs may be related to the protein's allergenicity [31]. To further develop these observations, the consideration of the propensity of various parvalbumins for intrinsic disorder was included in the analysis.

2. Materials and methods

2.1. The choice of parvalbumin sequences

The UniProt Knowledgebase (UniProtKB) [32] release 2013_03 (Mar 6, 2013) was searched for full-length parvalbumin and oncomodulin sequences, while PA-like and putative PA sequences were excluded (the query “(name:parvalbumin OR name:oncomodulin NOT name:like NOT name:putative) AND fragment:no” is used). The found 212 entries, including 72 Swiss-Prot entries, were exported into FASTA format. The entries P86431, P86432, P09227 with uncertain residues (labeled as ‘X’ in the FASTA file), have been removed. Alternative sequences and natural variants (searched using “Sequence annotation (ALTERNATIVE_SEQUENCE)” and “Sequence annotation (NATURAL_VARIANT)” fields, respectively) were found only for P02628 entry. The only additional natural variant of this entry was included into the FASTA file (P0262 – ID is assigned to its α_1 isoform [29]). Sequence conflicts were searched using “Sequence annotation (SEQUENCE_CONFLICT)” field. The corresponding previously reported PA sequences found for 7 entries (Q804W2, P80026, P19753, P0CE72, P20472, P32848, P02631) were added to the FASTA file (the last one or two symbols of UniProt ID were replaced by ‘-’). The resulting sequences were aligned using MUSCLE v.3.8.31 for Windows [33] with default settings, optimized for best accuracy. The aligned FASTA file was searched for the presence of problematic entries within the same organism (identical entries, truncated sequence versions, and the entries differing only by the presence of initiating N-terminal methionine) using Microhet v.1.0 software (IBI RAS, Pushchino, Russia). The following problematic entries have been removed from the FASTA file: Q1LWD7 (identical to Q804W2 entry), E1UIZ8 (identical to Q90YK9), C1L370 (differs from P80026 by initiating Met residue), B1AH72 (N-terminally truncated version of P2047- and P20472), Q545M7 (identical to P32848), Q8JIU1 (differs from P02617 by initiating Met), Q8JIU2 (differs from P02627 by initiating Met), Q3C2C3 (identical to P59747). Overall, 210 unique sequences of PA, including 78 Swiss-Prot sequences, were aligned using MUSCLE v.3.8.31. The alignment results were sorted by organism name using Microhet v.1.0 software (see supplementary file “Protein alignment.afa” in aligned FASTA format) and visualized using Geneious Pro 5.4.6 software [34] (see Fig. S1). The numbering of PA sequences is according to Fig. S1.

2.2. Analysis of parvalbumin sequences

Analysis of microheterogeneity within 210 unique PA sequences (see above) was performed using Microhet v.1.0 software (IBI RAS, Pushchino, Russia), written in Embarcadero® Delphi® 2010 v.14.0 (Embarcadero Technologies, Inc.) programming language. The pairwise sequence identity (SI) was defined as follows:

$$SI(\%) = 100 \cdot (\text{Number of identical positions}) / (\text{Number of overlapping positions}). \quad (1)$$

The pairs of protein isoforms with SI exceeding 90% are referred to as “microheterogeneous isoforms” of a protein. Each unique pair

of the microheterogeneous isoforms corresponds to a separate case of amino acid sequence microheterogeneity of a protein. The following types of analysis are supported in Microhet v.1.0 software: search of problematic entries (see above), calculation of *SI* values and analysis of their distributions, search of unique cases of microheterogeneity and calculation of statistical data on these cases, including distribution over aligned sequence of the unique substitutions giving rise to sequence microheterogeneity. Manual handling of the aligned sequences (comparison of the original and aligned scales of residues, etc.) was performed using Geneious Pro 5.4.6 software [34]. Some statistical data on microheterogeneous substitutions (the substitutions markedly affecting protein's disorder propensity and its stability, the residues located in species-specific and cross-reactive epitopes, etc.) were manually extracted from the data presented in Table 1.

2.3. Predictions of intrinsic disorder

Estimates of the per-residue propensity for intrinsic disorder in PAs were performed using the PONDR® VSL2P algorithm applicable to protein regions of any length [35], as implemented in PONDR analysis v.1 software (IBI RAS, Pushchino) [36]. Classification of amino acid substitutions by change in intrinsic disorder level was based on the scale of order-promoting abilities of residues described earlier [37]. The order-promoting residues were C, W, Y, I, F, V, L (set 1), the neutral residues were H, T, N, A, G, D (set 2), and the disorder-promoting residues were M, K, R, S, Q, P, E (set 3). The substitution was considered to exert drastic effect on protein's intrinsic disorder level when a conversion between sets 1 and 3 takes place. Any other conversion between differing sets of residues was supposed to exert only moderate effect.

Table 1

The list of unique cases of sequence microheterogeneity within 210 parvalbumin sequences. The numbering of sequences is according to Fig. S1. The sequences being allergens [1] are shown in bold. The residues affected by PA microheterogeneity are indicated, along with their involvement into known epitopic regions [42,43] (the residues located in the species-specific and cross-reactive epitopes are shown underlined and in bold, respectively) and the substitution-induced changes in intrinsic disorder propensity (the substitutions with mild and drastic changes are shown underlined and in bold, respectively) and in free energy of protein denaturation, $\Delta\Delta G$ (the substitutions with absolute value of $\Delta\Delta G$ above 0.5 kcal/mol are shown in bold). The estimates of intrinsic disorder changes are based on the scale of order-promoting abilities of residues described by Radivojac et al. [37] (for details, refer to 'Material and methods' chapter). The estimates of $\Delta\Delta G$ are performed using I-Mutant3.0 online service [38].

Sequence No. 1: Sequence number, UniProtKB ID_Organism name (lineage)	Sequence No. 2: Sequence number, UniProtKB ID_Organism name (lineage)	Sequence identity (%)	Index numbers of mutated residues using the scale of aligned residues (original numbering) Epitopic regions are marked	Amino acid substitutions, Sequence No. 1 → Sequence No. 2 Disorder changes are marked	$\Delta\Delta G$ (kcal/mol)
8, C1KUF7_BOVIN (α)	9, QOVCG3_BOVIN ^a (α)	95.46	<u>3 (2)</u> , <u>21 (18)</u> , 22 (19), 59 (55), 61 (57)	A → S, P → A, Y → F, Q → K, D → G	−0.65 , −1.44 , −0.06, −0.19, −0.77
18, E0WDA1_CLUHA (β)	19, C6GKU8_CLUHA (β)	99.08	97 (93)	V → D	−1.22
21, C6GKU7_CLUHA (β)	22, E0WDA6_CLUHA (β)	98.18	37 (33), 51 (47)	K → E, P → A	−0.28, −1.36
26, Q8UUS3_CYPCA (β)	27, E0WD92_CYPCA (β)	99.08	<u>23 (20)</u>	Q → K	−0.13
28, Q8UUS2_CYPCA (β)	29, P02618_CYPCA ^a (β)	97.22	15 (13/12), 33 (29/28), 78 (74/73)	T → A, T → A, G → D	0.785 , 0.57 , 0.23
28, Q8UUS2_CYPCA (β)	30, E0WD93_CYPCA (β)	98.17	15 (13), 33 (29)	T → A, T → A	−0.64 , −1.40
29, P02618_CYPCA ^a (β)	30, E0WD93_CYPCA (β)	99.07	78 (73/74)	D → G	−0.23
34, Q804W2_DANRE ^a (α)	35, Q804W-_DANRE ^a (α)	99.08	7 (6)	L → I	−0.98
39, Q804V8_DANRE (β)	40, Q6IMW7_DANRE (β)	97.25	<u>97 (93)</u> , 105 (101), 110 (106)	Y → D, Y → D, F → L	−0.99 , −1.10 , −0.73
48, P02628_ESOLU ^a (α)	49, P0262-_ESOLU ^a (α)	98.15	<u>16 (11–12/12)</u> , <u>33 (27/28)</u> , <u>37 (31/32)</u>	n.a. → L, <u>K → A</u> , L → K	n.a., 0.105, −1.755
58, Q90YK9_GADMO ^a (β)	59, A51874_GADMO (β)	99.08	45 (41)	P → S	−1.48
60, Q90YL0_GADMO (β)	61, A51873_GADMO (β)	99.08	105 (101)	D → E	−0.13
64, P80026_CHICK ^a (α)	65, P8002-_CHICK ^a (α)	96.33	66 (61), 104 (99), 105 (100), 106 (101)	<u>E → D</u> , A → V, D → E, E → K	−0.39, −0.16, 0.03, −0.66
66, P19753_CHICK ^a (β)	67, P1975-_CHICK ^a (β)	99.08	<u>100 (96)</u>	G → D	−0.73
73, P0CE72_HUMAN ^a (β)	74, P0CE7-_HUMAN ^a (β)	99.08	<u>23 (20)</u>	R → Q	−0.15
75, P2047-_HUMAN ^a (α)	76, P204-_HUMAN ^a (α)	99.07	<u>5 (4)</u> , 103 (99/98–99), 104 (100/98–99)	N → T, G → n.a., V → n.a.	0.14, n.a., n.a.
75, P2047-_HUMAN ^a (α)	77, P20472_HUMAN ^a (α)	99.09	<u>5 (4)</u>	N → T	0.87
76, P204-_HUMAN ^a (α)	77, P20472_HUMAN ^a (α)	100	<u>103 (98–99/99)</u> , 104 (98–99/100)	n.a. → G, n.a. → V	n.a., n.a.
88, Q5IRB2_LATCA (β)	89, Q6ITV0_LATCA (β)	99.08	76 (72)	S → F	0.48
105, P86745_MERAA ^a (β)	106, P86747_MERAA ^a (β)	99.07	<u>33 (28)</u>	V → E	−1.29
111, P56503_MERBI ^a (β)	112, P86753_MERBI ^a (β)	95.37	<u>14 (11)</u> , 41 (36) , 46 (41) , 107 (102), 109 (104)	V → I, T → S, D → E, W → F, <u>A → V</u>	0.08, −0.68 , −0.25, −1.01 , −0.02
119, P02620_MERME ^a (β)	120, P86765_MERME ^a (β)	99.07	105 (100)	E → D	−0.40
132, P32848_MOUSE ^a (α)	133, P3284-_MOUSE ^a (α)	99.09	114 (110)	S → T	−0.08
139, E0WDA2_ONCMY (β)	140, E0WDA4_ONCMY (β)	99.08	<u>81 (77)</u>	V → E	−1.05
146, P02624_RABIT ^a (α)	147, G1TF22_RABIT (α)	93.64	26 (22–23/23), 106 (102/103), 109 (105/106), 110 (106/107), 111 (107/108), 112 (108/109), 113 (109/110), 114 (110/111)	n.a. → A, E → A, T → N, L → F, V → T, S → Q, E → P, S → R	n.a., −0.83 , −0.97 , −1.10 , −1.50 , −0.3, −0.21, 0.13
161, P02631_RAT ^a (β)	162, P0263-_RAT ^a (β)	99.08	30 (26)	E → Q	−0.22
163, D3Z9U5_RAT (α)	164, P02625_RAT ^a (α)	95.46	106 (102), 107 (103), 110 (106), 112 (108), 114 (110)	V → E , H → F, I → L, G → A, P → S	−1.27 , 0.28, −0.43, −0.54 , −1.30
166, B9EMJ3_SALSA (β)	167, B9EPT7_SALSA (β)	99.08	106 (102)	V → E	−1.25
170, E0WD98_SALSA (β)	171, Q91482_SALSA ^a (β)	97.25	11 (9), 30 (26), 105 (101)	D → E, N → S, N → D	0.64 , −0.09, −0.17
173, Q91483_SALSA ^a (β)	174, E0WD99_SALSA (β)	97.22	<u>3 (2)</u> , 42 (37) , 45 (40)	<u>S → A</u> , <u>S → G</u> , S → N	−0.49, −0.83 , −0.45
183, D3GME4_SCOSC (β)	184, E0WD95_SCOSC (β)	98.17	88 (84), 109 (105)	K → E, <u>A → S</u>	−0.09, −0.70
186, C6GKU4_9PERC (β)	187, E0WD96_9PERC (β)	99.08	107 (103)	F → C	−1n26
191, B9VJM3_SINCH (β)	192, D2KQC3_SINCH (β)	99.08	<u>21 (18)</u>	A → V	−0.15

n.a., not available.

^a PA sequences from Swiss-Prot database.

within both PA lineages reveals that it is 2.0-fold higher for less allergenic α isoform: 50% versus 24% for β isoform.

Qualitatively identical results were obtained using Allergome Database v.4.0 (Mar 6, 2013; <http://www.allergome.org/>) [41]. Overall, 125 PAs are considered as allergens (7 and 118 for α and β PA lineages, respectively). 37 of them are microheterogeneous isoforms (2 and 35 for α and β isoforms, respectively). Hence, the occurrence of allergens is 1.05 times higher among PAs having microheterogeneous isoforms (incidence of allergens is 62% and 59% for microheterogeneous and non-microheterogeneous PA isoforms, respectively). This value is substantially lower than that derived from IUIS Allergen Nomenclature Sub-Committee web site, which reflects the tendency of Allergome Database to overestimation of the number of allergens.

It is shown using different databases of allergens that sequence microheterogeneity within PAs positively correlates with allergenicity rate among them: presence or higher level of sequence microheterogeneity is associated with higher allergenicity rate. Meanwhile, the incidence of sequence microheterogeneity within less allergenic α isoform exceeds that for more allergenic β isoform.

3.3. Analysis of amino acid substitutions accompanying sequence microheterogeneity in parvalbumin

Examination of the distribution over aligned PA sequence of 71 unique cases of amino acid substitutions (affect 39 residues) giving rise to sequence microheterogeneity (Table 1) reveals fairly uniform distribution of the substitutions, except for helices A, B, F and helices D and E, which correspond to the most and least affected regions, respectively (Fig. 2). The helices A, B, and F (33 residues) correspond to 25% of the aligned sequence length (132 residues), but comprise 59% of all the substitutions cases (42 substitutions). This region contains most of the residues which were substituted 3 and more times (7 from 8 residues). The fact that microheterogeneous substitutions in PA are abundant and concentrated in specific regions of the polypeptide chain, points out that these substitutions are unlikely to be a result of erroneous sequencing.

The per-residue disorder predictions using PONDR® VSL2P algorithm [35] identify the N-terminal part of the helix F and C-terminus of the EF-loop as regions with well-defined tendency to form an ordered structure (Fig. 2). Metal binding to the EF subdomain is expected to further stabilize these regions. The case-specific analysis of impact of the microheterogeneous substitutions on protein's intrinsic disorder level using the scale of order-promoting abilities of residues described earlier [37] has revealed (Table 1) that among 71 substitution cases 8% (6 substitutions involving 5 residues) is expected to be accompanied by the drastic changes in protein's disorder propensity, and 37% (26 substitutions involving 19 residues) should cause only moderate changes. Overall, 45% of the substitution cases are likely to lead to notable changes in protein disorder propensity, and respective changes in protein stability. The latter conclusion has been verified by calculations of free energy changes ($\Delta\Delta G$) accompanying the microheterogeneous substitutions using I-Mutant 3.0 online service [38] (Table 1). Assuming that $\Delta\Delta G$ value of 0.5 kcal/mol is used as a threshold for classification between neutral and stabilizing/destabilizing mutations [38], 48% of the microheterogeneous substitutions in Table 1 are classified as notably affecting protein stability.

Overlay of the distributions over aligned sequence of PA of the unique substitutions accompanying sequence microheterogeneity, known IgE-binding epitopes of β -PA from Atlantic salmon (*Salmo salar*) (identified by peptide-based microarray immunoassay [42]) and immunologically reactive fragments of cod (*Gadus callarias*) PA (derived by trypsinolysis [43]) is shown in Fig. 2 (see Table 1). 44% of these substitution cases (31 of 71 substitutions, affecting 20 residues) are located in the union of these species-specific epitopes (52% of protein length, 68 of 132 residues), and 11% (8 substitutions, affecting 6 residues) are located in the cross-reactive epitopes (17% of

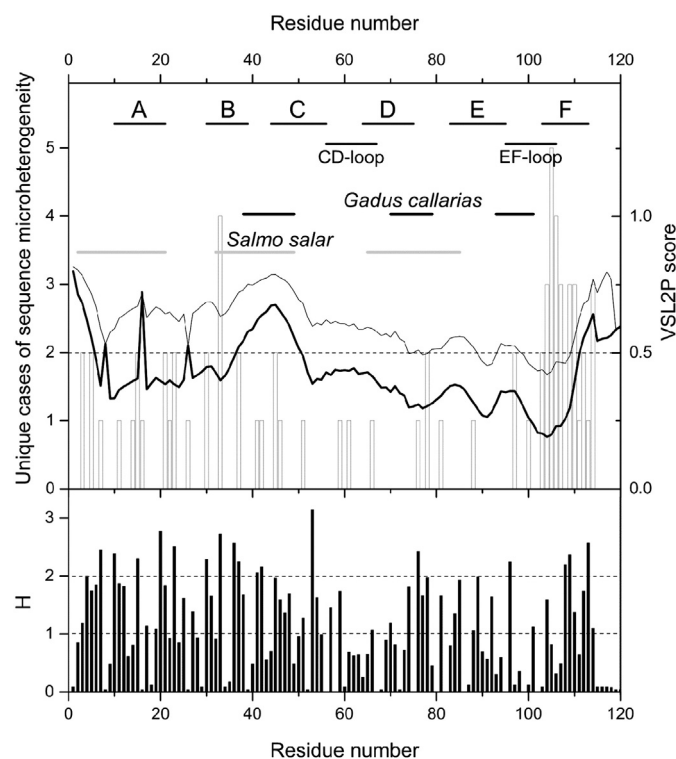


Fig. 2. Distribution over aligned sequence of PAs of the unique substitutions giving rise to sequence microheterogeneity (open bars) and Shannon diversity index [40] (H, solid bars). IgE-binding epitopes of PAs (bold lines) and average PONDR® VSL2P scores [35] (curves) are marked. The grey bold lines mark the IgE-binding regions of β -PA from Atlantic salmon (*Salmo salar*) [42]. The black bold lines indicate positions of the immunologically reactive fragments of cod (*Gadus callarias*) PA [43]. The thin and bold curves correspond to the highest and lowest estimates of VSL2P scores for PAs, respectively (standard deviation is either appended or subtracted from the average VSL2P curve). The dashed lines correspond to VSL2P score value 0.5 (the borderline between ordering and disordering) or Shannon diversity index values 1 and 2 (correspond to borderlines between different residue conservation levels). Positions of key secondary structure elements are indicated (taken from PDB entry 4CPV): helices from A to F, and Ca^{2+} -binding loops CD and EF.

protein length, 22 residues). Hence, the substitutions giving rise to sequence microheterogeneity within PA may affect IgE-binding through direct involvement in epitopic regions of PA. Another possibility is an allosteric effect of structural perturbations exerted by the mutations outside of these regions, since 34% of the substitutions are predicted to exert notable structural/energetic impact and located beyond the epitopic regions (see Table 1).

The spatial distribution of 39 residues engaged into sequence microheterogeneity of PA over X-ray structure of the Ca^{2+} -loaded state of intact carp parvalbumin pI 4.25 (PDB code 4CPV, UniProtKB entry P02618) is shown in Fig. 3. Examination of the tertiary structure using ASAView server [44] reveals that solvent accessibility of 49% of the 39 residues exceeds 50%. The overall surface distribution of the microheterogeneous substitutions is rather uniform, except for the region of helix F exhibiting higher substitution rates, and an extended cluster of residues 62–75 of the aligned scale (original residues 57–70), excluding residue 66. This cluster corresponding to the C-terminal half of CD-loop and helix D (Fig. 2) lacks microheterogeneous substitutions, but greatly intersects with IgE-binding epitopes of PAs from Atlantic salmon [42] (residues 65–85) and cod [43] (residues 70–79) (Figs. 2 and 3). Thus, this cluster includes an evolutionary conserved cross-reactive epitopic region 70–75 (C-terminus of helix D). Since the 70–75 region is a part of the only globally highly conserved stretch of residues 60–73 (see below), this region is likely to be of outstanding importance for the cross-reactivity of PAs.

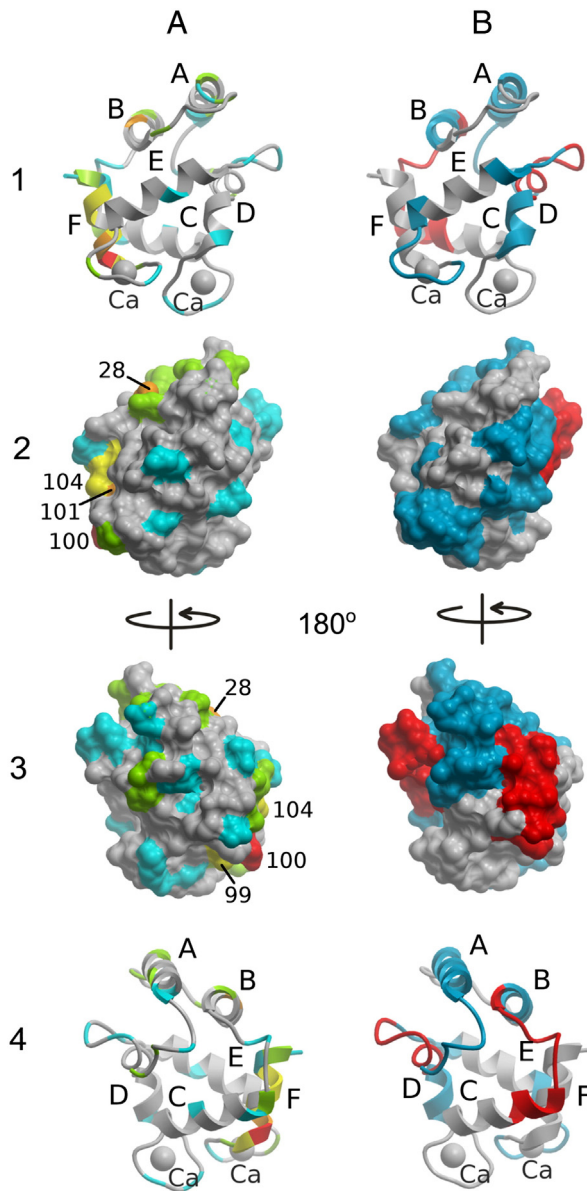


Fig. 3. The location of the residues affected by sequence microheterogeneity within 210 PAs (A) and the known IgE-binding epitopes of PAs from *Gadus callarias* and *Salmo salar* [42,43] (B) on the X-ray crystal structure of the Ca^{2+} -loaded form of intact carp parvalbumin pI 4.25 (PDB code 4CPV, UniProtKB entry P02618). (1) Ribbon diagram of the PA molecule, (2) molecular surface diagram of the view shown in panel (1), (3) 180° rotation of the view shown in panel (2), (4) ribbon diagram of the view shown in panel (3). The residues exhibiting 1, 2, 3, 4, 5 cases of sequence microheterogeneity are shown in cyan, green, yellow, orange, and red, respectively. The numbers of the three last categories of residues are indicated, using the scale of aligned residues. The residues corresponding to the species-specific and cross-reactive epitopes are colored in blue and red, respectively. Sequence alignment is performed using MUSCLE 3.8.31 algorithm [33]. The α -helices are labeled A–F. The figure was prepared with Accelrys Insight II software.

3.4. Evolutionary aspects of sequence microheterogeneity in parvalbumin

Since alternative splicing has not been reported for PA, sequence microheterogeneity reflects evolutionary processes occurring in PA genes, namely, either beginning of divergence of duplicate genes or appearance of allelic variants of the same gene. In any case, the distribution of microheterogeneous substitutions over PA structure represents a snapshot of recent evolutionary changes within the PA genome. Hence, the helices A, B and F, being the most actively changed regions due to the microheterogeneous substitutions (Fig. 2), can be regarded as evolutionary hotspots of PA. The direct involvement of helix F into

metal binding indicates that recent evolution of PA was aimed at modulation of the metal-binding properties of its EF subdomain. Since AB domains of pike α -PA and rat PA were previously shown to modulate their Ca^{2+} -binding properties [45,46], microheterogeneous substitutions in AB domain are likely aimed at alteration of the metal-binding properties of the protein.

Distribution of the Shannon diversity index (which is a measure of the sequence variability [40]) over aligned sequence of PAs (Fig. 2) reveals very similar distribution of evolutionary hot regions (AB domain and helix F) and highly conserved areas (helix D and C-terminal half of CD-loop) of the protein, which implies that direction of the recent PA evolution is similar to that of the overall evolution of the protein. Nevertheless, some peculiar features of the overall PA evolution are observed, including the variability of residues 53 and 96, located in the helix C and the EF-loop, respectively.

Examination of the PA genes from 9 species (including human, bovine, rabbit, guinea pig, rat, mouse, zebrafish, barramundi and rainbow trout) shows that breaks between the coding exons 2 to 5 correspond to residues 24, 69, 106 of the aligned residue scale, respectively. Since the last residue is located in the helix F in between the residues affected by microheterogeneity, recent evolution of the helix F was accompanied by the changes in both exons 4 and 5. The other breaks between coding exons (residues 24 and 69) are located beyond the actively evolving PA regions (Fig. 2). Another specific feature of the helix F is its proximity to the extremum of the PONDR® VSL2P score curve. Similar set of attributes (active evolutionary changes ensuring new functions at a break between coding exons, corresponding to extremum of PONDR® VSL2P score curve) has been previously observed in the hinge region of S100 proteins [36], which is intended for target recognition. Opposite to PA, PONDR® VSL2P score of the hinge region in S100 proteins reaches local maximum. Besides, helix F of PA and hinge region of S100 protein significantly differ in their location within the paired EF-hand motif. Despite these differences, the existence of regions with the very specific set of attributes in Ca^{2+} buffering PA and Ca^{2+} sensor S100 proteins indicates that these regions are inherent to both types of calcium-binding proteins.

Another peculiar region of PA sequence is the BC-loop and N-terminal half of the helix C, exhibiting marked tendency for intrinsic disorder and located in the cross-reactive epitope (Fig. 2). This region directly interacts with the helix F (Fig. 3), thereby contributing to fastening of the N- and C-termini of the paired EF-hand motif (that consists of CD and EF subdomains). The propensity for intrinsic disorder of the BC-loop and N-terminal part of helix C in PA seems to facilitate their interaction with the more ordered helix F, similarly to the interactions taking places during the process of S100 heterodimerization [36], in which structural rigidity of one of the interaction partners is compensated by structural mobility of another partner. The involvement of the helix F into the contacts stabilizing the paired EF-hand motif indicates that microheterogeneous substitutions in the helix F could affect structural stability of this motif. Furthermore, since the BC-loop and N-terminal half of helix C belong to the cross-reactive epitope, minor changes in the helix F may indirectly affect structural and dynamic properties of this epitopic region.

4. Discussion

The previously reported sequence microheterogeneity in pike α -PA [29] is shown here to be phenomenon inherent to PAs of at least 20 species, and hence it is much more common for PAs than it was expected earlier. It also expected that the further extension of our knowledge of PA sequences will show that sequence microheterogeneity is even more widespread among PAs. Since some of the microheterogeneous substitutions are concentrated within the physiologically significant PA regions (for example, helix F (Fig. 2)), it can be speculated that sequence microheterogeneity is a way of tuning the musculature

aimed at its adaptation to developmental and/or environmental conditions [29], as it has been proposed for different PA isoforms [27,28,47]. Considering that other muscle proteins (for example, some of actins [29]) also have microheterogeneous isoforms, it is likely that sequence microheterogeneity of muscle proteins is vital for ensuring the diversity of functions accomplished by various muscle types.

The data presented here evidence a well-defined regularity: the higher level of amino acid sequence microheterogeneity in PA is associated with its higher allergenicity. Hence, sequence microheterogeneity may serve as structural factor contributing to the allergenicity of PA. Meanwhile, an alternative view could be suggested. Sequence microheterogeneity may reflect active evolution of PA genes and respectively higher number of PA homologs with sequence identity below 90%, which could be even more important for allergenicity of PA. However, in some cases (for example, major cow's milk allergen Bos d 5, and major house dust mite allergens Der p 1, Der p 2 and Der f 1) [3] all protein isoforms are microheterogeneous. Hence, the presence of homologs of an allergen with sequence identity below 90% is not a prerequisite for its allergenicity.

It should be noted that the presented statistics on occurrence of allergens and microheterogeneous isoforms among different PA sets is likely to be affected by several factors. One of them is potential bias of research interests towards allergenic PAs, which may lead to overrepresentation of microheterogeneous isoforms of allergenic proteins with respect to non-allergenic PAs. The related potential issue is incomplete knowledge of PA isoforms. Besides, the information on levels of their expression is lacking. Furthermore, one should not exclude the presence of PCR or sequencing artifacts, serving as a source of erroneous microheterogeneous isoforms. The latter possibility can be partly ruled out since the microheterogeneous substitutions are fairly common for PAs and concentrated in specific regions of the protein (Fig. 2). Finally, allergenicity of PAs from non-commercial fish species remains unexplored. Considering the aforementioned factors, the statistics on the occurrence of allergens and microheterogeneous isoforms among different PA sets needs further rectification. Evidently, the conclusion on the relationship between sequence microheterogeneity in PA and its allergenicity derived from the bioinformatic analysis to be verified by detailed experimental studies of the immune response elicited by microheterogeneous isoforms of PAs, as exemplified by studies of other allergens [4–11].

Minor modifications of amino acid sequences of allergens were previously shown to exert multifaceted action on mechanisms of allergic sensitization and triggering of allergic reactions. For example, IgE and T cell reactivity patterns, as well as IgE and IgG responses can be affected [4–11]. From a statistical point of view, one may expect that microheterogeneous isoform of an allergen may induce either more positive or more negative overall immune response. Some of the microheterogeneous isoforms are likely to elicit more negative immune response than others, which might rationalize the increased allergenicity rate among microheterogeneous isoforms of PA. One of the possible ways of pro-allergic action of microheterogeneous substitutions within PA is altering of its IgE-binding via direct modification of the IgE-binding epitopes of PA (Fig. 2) or via allosteric mechanism, in which IgE-binding PA regions are affected by structural/energetic perturbations induced by distant substitutions (Table 1).

The potential of influence of microheterogeneity within PAs on immune response may be higher considering the possibility that PA isoforms from closely related species may have sequence identity exceeding 90%. Analysis of the 16 PAs established as allergens according to the WHO/IUIS Allergen Nomenclature Sub-committee [1] shows that 4 of them (sequences 60, 61, 171 and 201) have from 1 to 8 such homologs. The highest number of the homologs corresponds to sequences 60 (*Gadus morhua*) and 171 (*Salmo salar*) (8 and 5 homologs, respectively). Except for the sequences 11 and 38, these homologous PAs originate from the commercial fish species. The total number of the affected commercial fish species is 14. There is a possibility that consumption of these fish species is accompanied by the

cross-reactions with their PAs, in which PAs causing sensitization and triggering of allergic reactions originate from different species. The latter possibility can be achieved via numerous combinations, which likely increases the chances to cause allergic reaction.

Apparently, the phenomenon of sequence microheterogeneity alone cannot be responsible for the allergenicity of PAs, and other factors, for example, PA expression levels [21,26], high structural stability of calcium-loaded PA [25], and proximity of PA sequences to human homologs [24,48], are of importance. Meanwhile, the concept of sequence microheterogeneity enables us to put forward a hypothesis explaining the facts that allergy is more common for population of developed countries than developing countries, and higher rate of allergic diseases is observed in urban populations versus rural populations (for review, see [49]). The globalization processes inherent to developed countries and urban areas greatly diversified the ration of their population, which increases exposure to isoforms of food allergens, since some cultivars differ by allergen composition [50]. Hence, the population of developed countries is subjected to wider spectrum of allergen isoforms, including microheterogeneous isoforms, which likely to cause more frequent sensitization and triggering of allergic reactions. On the contrary, more traditional ration inherent to population of developing countries and rural areas decreases chances of contact with microheterogeneous isoforms of food allergens and respectively lowers the probability of both allergic sensitization and triggering of allergic reactions. This hypothesis apparently complements other existing explanations of these regularities (e.g., alterations in exposure to microorganisms and common parasites [49,51–53]).

Sequence microheterogeneity is not the only way of fine structural alterations of allergens which can modulate the immune response. Posttranslational modifications (e.g. phosphorylation) were previously shown to affect IgE response to caseins [54]. Furthermore, posttranslational modifications other than glycosylation were reported to convert self-proteins to the immunogenic variants (for review, see [55]). The residues S73, S79, T83 (77, 83, 87 of the aligned scale) of mouse α -PA (UniProtKB entry P32848) were reported to be phosphorylated [56]. These residues do not overlap with known sites of PA microheterogeneity (Table 1), but intersect with IgE-binding epitopes of β -PA from Atlantic salmon [42] (residues 77, 83 of the aligned scale) and cod PA [43] (residue 77). Hence, phosphorylation of these residues is potentially able to affect the IgE-binding. Furthermore, posttranslational modifications, including α -N-acetylation, seem to explain the differences in IgG and IgE binding capacities observed for recombinant and intact forms of pollack PAs [57]. This effect is expected, considering that similarly to the amino acid sequence microheterogeneity the posttranslational modifications of PA might enhance its structural microheterogeneity. Thus, the pro-allergic effect of the sequence microheterogeneity in PA is likely to be complemented by the analogous effect of its posttranslational modifications.

The validity and generality of the relationship between the sequence microheterogeneity and allergenicity of an allergen remain to be established. Apparently, other structural factors (for example, posttranslational modifications) which may increase structural microheterogeneity of an allergen should be taken into account. Meanwhile, in the case of generality of this relationship, the hygiene hypothesis explaining higher incidence of allergic diseases among population of developed countries and urban areas [49,51–53] will be extended by new considerations.

Acknowledgements

This work was supported by a grant from the Program of the Russian Academy of Sciences "Molecular and Cellular Biology" (S.E.P.), and a grant RFBR 11-04-00860-a from the Russian Foundation for Basic Research (S.E.P.). We are indebted to Prof. Eugene A. Permyakov, Dr. Alexei S. Kazakov and Victoria A. Rastrygina for careful reading and discussion of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbapap.2013.04.025>.

References

- [1] M.D. Chapman, A. Pomes, H. Breiteneder, F. Ferreira, Nomenclature and structural biology of allergens, *J. Allergy Clin. Immunol.* 119 (2007) 414–420.
- [2] M.F. Schenk, L.J. Gilissen, G.D. Esselink, M.J. Smulders, Seven different genes encode a diverse mixture of isoforms of Bet v 1, the major birch pollen allergen, *BMC Genomics* 7 (2006) 168.
- [3] S. Piboonpocanun, N. Malainual, O. Jirapongsananuruk, P. Vichyanond, W.R. Thomas, Genetic polymorphisms of major house dust mite allergens, *Clin. Exp. Allergy* 36 (2006) 510–516.
- [4] F. Ferreira, K. Hirtenlehner, A. Jilek, J. Godnik-Cvar, H. Breiteneder, R. Grimm, K. Hoffmann-Sommergruber, O. Scheiner, D. Kraft, M. Breitenbach, H.J. Rheinberger, C. Ebner, Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy, *J. Exp. Med.* 183 (1996) 599–609.
- [5] S. Wagner, C. Radauer, M. Bublin, K. Hoffmann-Sommergruber, T. Kopp, E.K. Greisenegger, L. Vogel, S. Vieths, O. Scheiner, H. Breiteneder, Naturally occurring hypoallergenic Bet v 1 isoforms fail to induce IgE responses in individuals with birch pollen allergy, *J. Allergy Clin. Immunol.* 121 (2008) 246–252.
- [6] H. Breiteneder, F. Ferreira, K. Hoffmann-Sommergruber, C. Ebner, M. Breitenbach, H. Rumpold, D. Kraft, O. Scheiner, Four recombinant isoforms of Cor a 1, the major allergen of hazel pollen, show different IgE-binding properties, *Eur. J. Biochem.* 212 (1993) 355–362.
- [7] S. Schenk, K. Hoffmann-Sommergruber, H. Breiteneder, F. Ferreira, G. Fischer, O. Scheiner, D. Kraft, C. Ebner, Four recombinant isoforms of Cor a 1, the major allergen of hazel pollen, show different reactivities with allergen-specific T-lymphocyte clones, *Eur. J. Biochem.* 224 (1994) 717–722.
- [8] B.J. Hales, L.A. Hazell, W. Smith, W.R. Thomas, Genetic variation of Der p 2 allergens: effects on T cell responses and immunoglobulin E binding, *Clin. Exp. Allergy* 32 (2002) 1461–1467.
- [9] J.W. Park, K.S. Kim, H.S. Jin, C.W. Kim, D.B. Kang, S.Y. Choi, T.S. Yong, S.H. Oh, C.S. Hong, Der p 2 isoallergens have different allergenicity, and quantification with 2-site ELISA using monoclonal antibodies is influenced by the isoallergens, *Clin. Exp. Allergy* 32 (2002) 1042–1047.
- [10] E.M. Gonzalez, M. Villalba, M. Lombardero, M. Aalbers, R. van Ree, R. Rodriguez, Influence of the 3D-conformation, glycan component and microheterogeneity on the epitope structure of Ole e 1, the major olive allergen. Use of recombinant isoforms and specific monoclonal antibodies as immunological tools, *Mol. Immunol.* 39 (2002) 93–101.
- [11] G. Gäfvelin, S. Parmley, T. Neimert-Andersson, U. Blank, T.L. Eriksson, M. van Hage, J. Punnonen, Hypoallergens for allergen-specific immunotherapy by directed molecular evolution of mite group 2 allergens, *J. Biol. Chem.* 282 (2007) 3778–3787.
- [12] C.H. Schein, O. Ivancic, T. Midoro-Horiuti, R.M. Goldblum, W. Braun, An allergen portrait gallery: representative structures and an overview of IgE binding surfaces, *Bioinform. Biol. Insights* 4 (2010) 113–125.
- [13] E.A. Permyakov, Parvalbumin, Nova Science Publishers, Inc., New York, 2006.
- [14] S.H. Arif, A Ca(2+)–binding protein with numerous roles and uses: parvalbumin in molecular biology and physiology, *BioEssays* 31 (2009) 410–421.
- [15] H. Belge, P. Gailly, B. Schwaller, J. Löffing, H. Debaix, E. Riveira-Munoz, R. Beauwens, J.P. Devogelaer, J.G. Hoenderop, R.J. Bindels, O. Devuyst, Renal expression of parvalbumin is critical for NaCl handling and response to diuretics, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 14849–14854.
- [16] S. Nakayama, N.D. Moncrief, R.H. Kretsinger, Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories, *J. Mol. Evol.* 34 (1992) 416–448.
- [17] C.E. Nockolds, R.H. Kretsinger, C.J. Coffee, R.A. Bradshaw, Structure of a calcium-binding carp myogen, *Proc. Natl. Acad. Sci. U. S. A.* 69 (1972) 581–584.
- [18] A. Bugajska-Schretter, L. Elfman, T. Fuchs, S. Kapiotis, H. Rumpold, R. Valenta, S. Spitzauer, Parvalbumin, a cross-reactive fish allergen, contains IgE-binding epitopes sensitive to periodate treatment and Ca²⁺ depletion, *J. Allergy Clin. Immunol.* 101 (1998) 67–74.
- [19] T. Van Do, S. Elsayed, E. Florvaag, I. Hordvik, C. Endresen, Allergy to fish parvalbumins: studies on the cross-reactivity of allergens from 9 commonly consumed fish, *J. Allergy Clin. Immunol.* 116 (2005) 1314–1320.
- [20] S.L. Taylor, J.L. Kabourek, S.L. Hefle, Fish allergy: fish and products thereof, *J. Food Sci.* 69 (2004) 175–180.
- [21] U. Griesmeier, S. Vazquez-Cortes, M. Bublin, C. Radauer, Y. Ma, P. Briza, M. Fernandez-Rivas, H. Breiteneder, Expression levels of parvalbumins determine allergenicity of fish species, *Allergy* 65 (2010) 191–198.
- [22] K. Aas, S.M. Elsayed, Characterization of a major allergen (cod). Effect of enzymic hydrolysis on the allergenic activity, *J. Allergy* 44 (1969) 333–343.
- [23] P. Hajeb, J. Selamat, A contemporary review of seafood allergy, *Clin. Rev. Allergy Immunol.* 42 (2012) 365–385.
- [24] J.A. Jenkins, H. Breiteneder, E.N. Mills, Evolutionary distance from human homologs reflects allergenicity of animal food proteins, *J. Allergy Clin. Immunol.* 120 (2007) 1399–1405.
- [25] A. Bugajska-Schretter, M. Grote, L. Vangelista, P. Valent, W.R. Sperr, H. Rumpold, A. Pastore, R. Reichelt, R. Valenta, S. Spitzauer, Purification, biochemical, and immunological characterisation of a major food allergen: different immunoglobulin E recognition of the apo- and calcium-bound forms of carp parvalbumin, *Gut* 46 (2000) 661–669.
- [26] A. Kobayashi, H. Tanaka, Y. Hamada, S. Ishizaki, Y. Nagashima, K. Shiomi, Comparison of allergenicity and allergens between fish white and dark muscles, *Allergy* 61 (2006) 357–363.
- [27] F. Hurliaux, P. Vandewalle, B. Focant, Immunological study of muscle parvalbumin isotypes in three African catfish during development, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 132 (2002) 579–584.
- [28] P. Brownridge, L.V. de Mello, M. Peters, L. McLean, A. Claydon, A.R. Cossins, P.D. Whitfield, I.S. Young, Regional variation in parvalbumin isoform expression correlates with muscle performance in common carp (*Cyprinus carpio*), *J. Exp. Biol.* 212 (2009) 184–193.
- [29] S.E. Permyakov, A.S. Karnoup, A.G. Bakunts, E.A. Permyakov, Sequence microheterogeneity of parvalbumin pl 5.0 of pike: a mass spectrometric study, *Biochim. Biophys. Acta* 1794 (2009) 129–136.
- [30] S.E. Permyakov, A.G. Bakunts, A.I. Denesyuk, E.L. Knyazeva, V.N. Uversky, E.A. Permyakov, Apo-parvalbumin as an intrinsically disordered protein, *Proteins* 72 (2008) 822–836.
- [31] B. Xue, D. Soeria-Atmadja, M.G. Gustafsson, U. Hammerling, A.K. Dunker, V.N. Uversky, Abundance and functional roles of intrinsic disorder in allergenic proteins and allergen representative peptides, *Proteins* 79 (2011) 2595–2606.
- [32] T.U. Consortium, Reorganizing the protein space at the Universal Protein Resource (UniProt), *Nucleic Acids Res.* 40 (2012) D71–D75.
- [33] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797.
- [34] A.J. Drummond, B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T.W.A. Thierer, Geneious 5.4 (2011).
- [35] K. Peng, P. Radivojac, S. Vucetic, A.K. Dunker, Z. Obradovic, Length-dependent prediction of protein intrinsic disorder, *BMC Bioinform.* 7 (2006) 208.
- [36] S.E. Permyakov, R.G. Ismailov, B. Xue, A.I. Denesyuk, V.N. Uversky, E.A. Permyakov, Intrinsic disorder in S100 proteins, *Mol. Biosyst.* 7 (2011) 2164–2180.
- [37] P. Radivojac, L.M. Iakoucheva, C.J. Oldfield, Z. Obradovic, V.N. Uversky, A.K. Dunker, Intrinsic disorder and functional proteomics, *Biophys. J.* 92 (2007) 1439–1456.
- [38] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.* 33 (2005) W306–W310.
- [39] S. Khan, M. Vihinen, Performance of protein stability predictors, *Hum. Mutat.* 31 (2010) 675–684.
- [40] M. Garcia-Boronat, C.M. Diez-Rivero, E.L. Reinherz, P.A. Reche, PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery, *Nucleic Acids Res.* 36 (2008) W35–W41.
- [41] A. Mari, C. Rasi, P. Palazzo, E. Scala, Allergen databases: current status and perspectives, *Curr. Allergy Asthma Rep.* 9 (2009) 376–383.
- [42] M. Perez-Gordo, J. Lin, L. Bardina, C. Pastor-Vargas, B. Cases, F. Vivanco, J. Cuesta-Herranz, H.A. Sampson, Epitope mapping of Atlantic salmon major allergen by peptide microarray immunoassay, *Int. Arch. Allergy Immunol.* 157 (2012) 31–40.
- [43] S. Elsayed, J. Apold, Immunochemical analysis of cod fish allergen M: locations of the immunoglobulin binding sites as demonstrated by the native and synthetic peptides, *Allergy* 38 (1983) 449–459.
- [44] S. Ahmad, M. Gromiha, H. Fawareh, A. Sarai, ASAVIEW: database and tool for solvent accessibility representation in proteins, *BMC Bioinform.* 5 (2004) 51.
- [45] E.A. Permyakov, V.N. Medvedkin, Y.V. Mitin, R.H. Kretsinger, Noncovalent complex between domain AB and domains CD*EF of parvalbumin, *Biochim. Biophys. Acta* 1076 (1991) 67–70.
- [46] M.T. Henzl, S. Agah, J.D. Larson, Association of the AB and CD–EF domains from rat alpha- and beta-parvalbumin, *Biochemistry* 43 (2004) 10906–10917.
- [47] T. Hasebe, K. Umezawa, M. Sugita, T. Iwata, K. Yamamoto, T. Obinata, S. Kikuyama, Postmetamorphic changes in parvalbumin expression in the hindlimb skeletal muscle of the bullfrog, *Rana catesbeiana*, *Biochim. Biophys. Acta* 1646 (2003) 42–48.
- [48] C. Santiago Hda, S. Bennuru, J.M. Ribeiro, T.B. Nutman, Structural differences between human proteins and aero- and microbial allergens define allergenicity, *PLoS One* 7 (2012) e40552.
- [49] M. Yazdanbakhsh, P.G. Kremsner, R. van Ree, Allergy, parasites, and the hygiene hypothesis, *Science* 296 (2002) 490–494.
- [50] C. Esteve, C. Montealegre, M.L. Marina, M.C. Garcia, Analysis of olive allergens, *Talanta* 92 (2012) 1–14.
- [51] D.P. Strachan, Hay fever, hygiene, and household size, *BMJ* 299 (1989) 1259–1260.
- [52] E.M. Carvalho, L.S. Bastos, M.I. Araujo, Worms and allergy, *Parasite Immunol.* 28 (2006) 525–534.
- [53] A.B. Fishbein, R.L. Fuleihan, The hygiene hypothesis revisited: does exposure to infectious agents protect us from allergy? *Curr. Opin. Pediatr.* 24 (2012) 98–102.
- [54] H. Bernard, H. Meisel, C. Creminon, J.M. Wal, Post-translational phosphorylation affects the IgE binding capacity of caseins, *FEBS Lett.* 467 (2000) 239–244.
- [55] T. Matsuda, T. Matsubara, S. Hino, Immunogenic and allergenic potentials of natural and recombinant innocuous proteins, *J. Biosci. Bioeng.* 101 (2006) 203–211.
- [56] J. Dai, W.H. Jin, Q.H. Sheng, C.H. Shieh, J.R. Wu, R. Zeng, Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry, *J. Proteome Res.* 6 (2007) 250–262.
- [57] T. Van Do, I. Hordvik, C. Endresen, S. Elsayed, Characterization of parvalbumin, the major allergen in Alaska pollack, and comparison with codfish allergen M, *Mol. Immunol.* 42 (2005) 345–353.
- [58] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics* 25 (2009) 1189–1191.