Journal of Bioinformatics and Computational Biology
Vol. 4, No. 2 (2006) 389–402

Imperial College Press
www.icpress.co.uk

SELECTION OF CRYPTIC B-CELL EPITOPES USING INFORMATIONAL ANALYSIS OF PROTEIN SEQUENSES

ELENA SVIRSHCHEVSKAYA

Department of Immunology
Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry RAS
Miklukho-Maklaya St., 16/10, Moscow, 117997, Russia
esvir@mail.ibch.ru

LUDMILA ALEKSEEVA

Laboratory of Peptide Chemistry
Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry RAS
Miklukho-Maklaya St., 16/10, Moscow, 117997, Russia
alexeeva@ibch.ru

ALEXEI MARCHENKO and SERGEI BENEVOLENSKII

Laboratory of Biotechnology of Eukaryotic Microorganisms
State Research Institute for Genetics and Selection of Industrial Microorganisms
1st Dorozniy proezd, Moscow, 113545, Russia

VALENTINA M. BERZHEC

Ilya Mechnikov Institute for Vaccine and Sera
Malyi Kasenniy per. 5A, Moscow, 105064, Russia

ALEXEI NEKRASOV

Group of Computer Graphic
Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry RAS
Miklukho-Maklaya St., 16/10, Moscow, 117997, Russia
alexei_nekrasov@mail.ru

Received 7 October 2005
Revised 6 December 2005
Accepted 8 December 2005

Sub-unit vaccines are synthetic or recombinant peptides representing T- or B-cell epitopes of major protein antigens from a particular pathogen. Epitope selection requires the synthesis of peptides that overlap the protein sequences and screening for the most effective ones. In this study a new method of immunogenic peptide selection based on the analysis of information structure of protein sequences is suggested. The analysis of known B-cell epitope location in the information structure of *Aspergillus fumigatus* proteins Asp f 2 and Asp f 3 has shown that epitopes are scattered along the sequences of proteins for the exception of sites with Increased Degree Information Coordination (IDIC). Based on these results peptides from different allergens such as Asp f 2, Der p 1, and Fel d 1 were selected and produced in a recombinant form in the context of yeast

virus-like particles (VLPs). Immunization of mice with VLPs containing peptides form allergens has induced the production of IgG able to recognize full-length antigens. This result suggests that the analysis of information structure of proteins can be used for the selection of peptides possessing cryptic B-cell epitope activity.

Keywords: B-cell epitopes; information structure of proteins; virus-like particles.

1. Introduction

Peptides corresponding to epitopes of proteins recognized by T- or B-cells are the best candidates for the sub-unit vaccines due to their hypoallergenic properties and low cost in comparison with full-length proteins. Epitope selection requires the synthesis of peptides that overlap the protein sequences and screening for the most effective ones. Considerable efforts have been made to develop computational tools for the analysis of protein sequences for the prediction of T- and B-cell epitopes in recent years. Several methods are currently available for the prediction of possible B-cell epitopes based on physicochemical or statistical characteristics of polypeptide chain segments.[1–4] The predictive value of different approaches varies significantly depending on the method of the prediction and the sequence of the protein.[5] B-cell epitopes, from five to nine amino acid (aa) long peptides recognized directly by immunoglobulins (Ig), can be defined as linear, which are formed by adjacent aa, and conformational, which are created during protein folding. Besides, some peptides represent so-called cryptic epitopes, which are able to mount IgG response when injected into mice in adjuvant however are not recognized when mice are immunized with full-length proteins containing these peptides.[6] Both IgG to natural or cryptic epitopes bind full-length proteins and thus can be protective.

Recently, a novel approach to the analysis of protein sequences permitting characterization of their information structure (IS) was suggested.[7–9] IS of proteins has a hierarchic organization and can be presented as IDIC-branches forming IDIC-trees, describing the location in the sequences of Increased Degree Information Coordination (IDIC) sites. Anomalously Distributed Density (ADD) of IDIC-branches in the lowest hierarchic level is likely to play an important role in the functional properties of proteins. It has been shown that areas characterized by the increased density of IDIC-branches (ADD^+-sites) are often found in sites responsible for enzymatic activity of proteins (personal communication). It can be hypothesized that peptides localized in different IS sites of proteins can possess different immunogenic properties. To analyze this hypothesis we have compared the localization of B-cell epitopes from *Aspergillus fumigatus* (Af) proteins Asp f 2 and Asp f 3 and IS of corresponding proteins. According to the results obtained six peptides from three different proteins Asp f 2, Der p 1, and Fel d 1, representing major allergens from Af, house dust mites and cat dander accordingly, were selected and produced in a recombinant form as yeast virus-like particles (VLPs). VLPs were used both as a peptide carrier and the source of T-cell epitopes. It is known that T-cell help is essential for switching B-cells to produce IgG.[10] VLP is often used for antigen delivery and is exploited in the vaccines against human immunodeficiency virus

(HIV) and the malarias.[11,12] Earlier we have shown that VLP can also be utilized as a vehicle for peptide delivery for the purposes of allergic response amelioration.[13] Immunogenic properties of predicted peptides were studied.

2. Materials and Methods

2.1. *Animals*

Ten to 12 weeks old inbred BALB/c (H-2^d), CBA (H-2^k) or C57BL/6 (H-2^b) mice bred at the Institute of Bioorganic Chemistry Animal Department, Moscow, were used in this study. The Institutional Animal Use and Care Committee approved the experimental protocol followed in this research.

2.2. *Antigens*

Recombinant Asp f 2 and Asp f 3 from Af were cloned and expressed in *Escherichia coli* using a pET vector.[14] The "C" terminal histidine tag was used to purify the allergen by Ni-affinity chromatography. House dust mite (HDM) and cat dander extracts were purchased at the Institute of Vaccines and Sera, Moscow. Twelve peptides from Asp f 2 and 11 peptides from Asp f 3 were synthesized commercially (Alpha Diagnostic Int., San Antonio).

2.3. *In vivo experiments*

2.3.1. *B-cell epitope analysis*

Mice (five to seven mice per group) were immunized once in the paws with 50 μg of full-length Asp f 2 or Asp f 3 in complete Freund's adjuvant (CFA). Animals were bled from tail 20 days later and their sera were collected and stored under refrigeration until tested.

2.3.2. *Immunogenic properties of peptides selected by information structure analysis*

BALB/c mice (five per group) were immunized subcutaneously at the tail base with 100 μg of VLP peptides in 50 μl of phosphate buffer on three consecutive days. Sera were collected as described above on day 25 after the first immunization.

2.4. *Enzyme Linked Immunosorbent Assay (ELISA)*

Antigen specific IgGs were determined according to the instructions of the manufacturer. In brief, 100 μl of Asp f 2 and Asp f 3 (5 μg/ml of PBS) or crude extracts, (100 μg/ml of PBS) per well were coated on ELISA plates (Costar). All dilutions of the reagents were made in 1% of BSA in PBS (BSA-PBS) and incubations were made at room temperature. Plates were washed three times between incubations with 0.05% Tween 20 in PBS (PBS-T). Biotinylated IgG1 specific anti-mouse

conjugates (Pharmingen) and extravidin–HRP conjugate (Sigma Chemicals) were then added and incubated as before. The color was developed using orthophenylene diamine (OPD) as the substrate, and the optical density (O.D.) was measured at 492 nm on an ELISA reader. The data presented as net O.D. values after subtracting the blanks. B cell epitopes of Asp f 2 and Asp f 3 were determined by peptide ELISA. Plates were coated overnight with 5 μg/ml of synthetic peptides dissolved in carbonate buffer, pH 9.0, in 100 μl/well. For peptide ELISA all dilutions of the reagents were made in PBS-T. The remaining steps of the ELISA were performed as described above.

2.5. *Information Structure (IS) analysis*

IS approach is based on two specific features of information entropy identified earlier during the analysis of non-homologous protein data sets.[9]

(i) Dependence of information entropy on the distance between residues has an S–shaped form. This dependence points out to the existence of sites of different size with increased degree of information coordination (IDIC-sites).
(ii) The level of coordination between residues separated by six or less positions in proteins is constant and high. Thus, it can be postulated that the distance in six residues is highly coordinated, and this is true for all proteins. This interval in six residues was called Hight Level Coordination Distance (HLCD). It permits the consideration of a protein sequence by means of short fragments no longer than the HLCD — "information units" (IU). The description of a protein sequences in IU permits investigation of IS organization in more detail. To develop a graphical presentation of IS the next algorithm was used.
(iii) All protein sequence was coded in IU terms. Short overlapped sequence fragments of a fixed size were considered as IUs.
(iv) "Information content" profile of the protein sequence was generated using parameters of residue frequencies obtained earlier.[8,9] Characteristics of residue frequencies were analyzed earlier for three different sets of non-homologous protein data sets.[8,9]
(v) To determine IDIC-sites the "information content" profile of the protein sequence, obtained at the step 2, was decomposed into Gaussian curves with a fixed half-width. Position centers of Gaussian functions with the greatest values determine the centers of IDIC-sites. Computations with different values of Gaussian functions' half-width permit identification of IDIC-sites of different sizes.
(vi) The centers of IDIC-sites of different half-width represent information structure (IS) of the protein. IDIC-diagrams, the most clear representation of IS, can be constructed when the centers of IDIC-sites of different sizes are connected. When this is fulfilled hierarchically organized IDIC-branches and IDIC-trees are formed.

Closely located IDIC-branches of low hierarchic level are of a special interest in IS. Close location was determined as being 30% less than HLCD. These closely located IDIC-branches of low hierarchic level represent the sites in IS with Anomalously Distributed Density (ADD+). Corresponding sites in the protein sequence were called ADD^+-sites. In this work IDIC-branches forming ADD^+-sites in the protein sequences studied are shown in black (Figs. 1 and 3). IS analysis is described in more detail in Ref. 7. IS analysis was conducted for the sequences of Asp f 2, Asp f 3, Der p 1, and Fel d 1.

2.6. *Virus-like particles*

Peptides from Asp f 2, Der p 1, and Fel d 1 selected on the basis of IS analysis (Table 1) were expressed in the yeast *Saccharomyces cerevisiae* as virus-like particles (VLPs) as was reported earlier.[15] Double-stranded DNA fragments encoding peptides were obtained by annealing the synthetic oligonucleotides shown in Table 1. VLPs without heterologous inserts were used as a control preparation (VLP0). For the *in vivo* experiments VLP containing solutions were dialyzed against PBS and 0.2 μm filter sterilized before injecting animals.

2.7. *Statistics*

The means and standard deviations were calculated for each group, and the groups were compared by paired t-test analysis using Excell Statistic Program.

Table 1. Sequences of synthetic oligonucleotides for the production of recombinant peptides from Asp f 2, Der p 1, and Fel d 1.

Peptides	Pairs of Forward and Reverse Primers	Constructs
Der p 1 (97–112)	gatctgttgctagagaacaatcttgtagaagaccaaatgctcaaagatttggtg gatccaccaaatctttgagcatttggtcttctacaagattgttctctagcaaca	Dp-97–112
Der p 1 (145–162)	gatctaaagatttagatgcattcagacattatgatggtagaacaattattcaaagagatg gatccatctctttgaataattgttctaccatcataatgtctgaaagcatctaaatcttta	Dp-145–162
Fel d 1 (17–34)	gatctactcctgatgaatatgttgaacaagttgctcaatacaaagccttaccagttgttg gatccaacaacaggtaaggctttgtattgagcaacttgttcaacatattcatcaggagta	Fel-17–37
Fel d 1 (44–63)	gatcttgtgttgatgcaaaaatgactgaagaagataaggaaaatgctttgtctgttg gatccaacagacaaagcattttccttatcttcttcagtcatttttgcatcaataacacaa	Fel-44–63
Asp f 2 (60–71)	gatctatggaagcagttggtgcatatgatgtaatagtaaatg gatccatttactattacatcatatgcaccaactgcttccata	Af-60–71
Asp f 2 (235–249)	gatcttctggttcaggtgctactactactccaactgattcaccatcag gatcctgatggtgaatcagttggagtagtagtagcacctgaaccagaa	Af-235–249
Asp f 2[a] (254–268)	gatctccatccaattgtcatacacatgaaggtggtcagctgcattgtactg gatccagtacaatgcagctgaccaccttcatgtgtatgacaattggatgga	Af-254–268

[a]B-cell epitope of Asp f 2.

3. Results

3.1. *B-cell epitopes of Asp f 2 and Asp f 3*

In order to obtain the most reliable data, B-cell epitopes of Asp f 2 and Asp f 3 were determined using three strains of mice. Mice were immunized only once to prevent minor epitope recognition. The results of B-cell epitope analysis for different strains of mice are summarized in Tables 2–3. The antigenicity of the B-cell epitopes have

Table 2. IgG B-cell epitopes recognized by sera from BALB/c, CBA and C57BL/6 mice immunized with full-length protein Asp f 2.

			O.D.		
Peptide	Position	Peptide Sequences	CBA	BALB/c	C57BL/6
1	3–24[a]	GAVTSFPIHSSCNATQRRQIEA	1,20		
2	31–42	ELARHAKAHILR			1.29
3	47–58	SEIYRKYFGNRP		1.09[b]	
4	60–71	MEAVGAYDVIVN			
5	92–106	GWGGHWRGANATSET		**1.14**	**0.9**
6	113–127	YTTRRWLVSMCSQGY			
7	**140–154**	**SDLMHRLYHVPAVGQ**		**1.57**	**1.38**
8	**155–169**	**GWVDHFADGYDEVIA**	**0,69**[c]	**0.92**	**0.79**
9	172–186	KSNGTESTHDSEAFE			
10	212–226	DQGHDTGSASAPAST			
11	235–249	SGSGATTTPTDSPSA			
12	**254–268**	**PSNCHTHEGGQLHCT**	**1,83**	**1.83**	**1.71**
Asp f 2			2,91	2.48	2.65

[a]Synthetic peptides corresponding to the indicated position in the sequence of Asp f 2 were used to coat ELISA plates.
[b]Net values of optical density are shown. Sera pooled from five mice in each group were diluted 1:500. Standard deviation was less than 10% in all cases. Three separate experiments were performed and data from a representative experiment are shown.
[c]Dominant linear IgG B-cell epitopes recognized by at least two strains are shown in bold.

Table 3. IgG B-cell epitopes recognized by sera from BALB/c and C57BL/6 mice immunized with full-length protein Asp f 3.

			O.D.	
Peptide	Position	Peptide Sequences	BALB/c	C57BL/6
1	**1–15**[a]	**MSGLKAGDSFPSDVV**	1.03	1.75
2	16–30	FSYIPWSEDKGEITA		0.89
3	34–46	PINYNASKEWADK		
4	47–61	KVILFALPGAFTPVC		
5	68–82	EYIEKLPEIRAKGVD	0.91	
6	83–97	VVAVLAYNDAYVMSA	0.85	0.93
7	**97–111**	**AWGKANQVTGDDILF**	**1.61**	**1.07**
8	115–126	PDARFSKSIGWA		
9	124–138	**GWADEEGRTKRYALV**		0.96
10	143–156	KITYAALEPAKNHL		
11	154–168	NHLEFSSAETVLKHL		
Asp f 3			2.63	2.59

[a]See notes under the Table 2.

varied in different strains, although some epitopes (dominant) were recognized by all three mice strains studied.

Earlier we have shown that human IgG from allergic bronchopulmonary aspergillosis (ABPA) patients invariably recognize most peptides. It can be a result of either cross reactivity or huge amount of low affinity IgG to these proteins. We have studied only B-cell epitopes recognized by IgE from ABPA sera patients. The epitopes identified in earlier studies were as follows: 16–22, 51–55, 96–98, 113–117, 140–144, 147–150, 158–162, 238–244, 258–261 for Asp f 2[16] and 10–19, 29–34, 46–52, 70–75, 98–102, 120–130 and 140–150 for Asp f 3.[17]

3.2. *Comparison of B-cell epitope location and information structure of Asp f 2 and Asp f 3*

The structure of the information of protein sequences determined as described earlier has a hierarchic form and can be presented as IDIC-trees, describing the location and interdependence of IDIC-sites in the proteins.[7] IDIC-diagrams for Asp f 2 and Asp f 3 are shown in Fig. 1. These results are also shown in Fig. 2, where the centers of IDIC branches are shown in bold. The distance between IDIC branches in 1–4 residues corresponds to ADD$^+$-sites, which are labeled in black in Fig. 1 and shaded in gray in Fig. 2. B-cell epitopes recognized either by mice or humans are delineated (Fig. 2).

Several epitopes in some cases are delineated in one fragment as some murine and human epitopes overlapped. So, for the estimation we have considered every peptide recognized by any strain of mice or by human IgE as a separate epitope. Totally Asp f 2 contained 24 and Asp f 3 – 17 epitopes. The analysis of B-cell epitope location has shown that among 24 Asp f 2 B-cell epitopes 80% have located

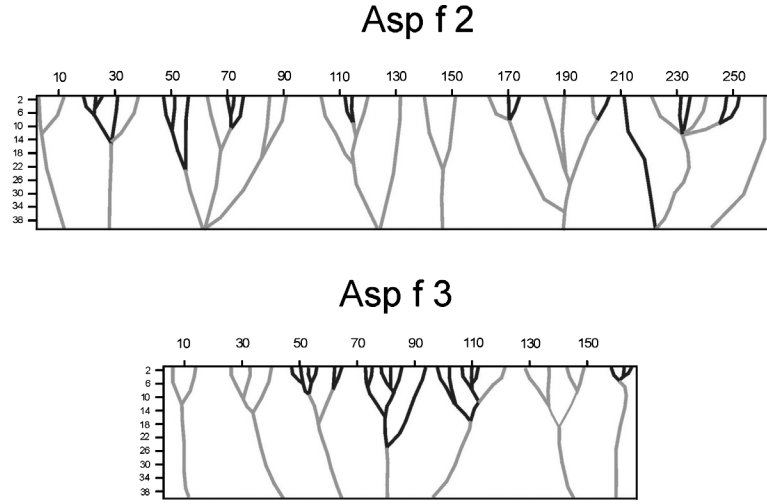

Fig. 1. IDIC-diagrams of Asp f 2 and Asp f 3. IDIC-branches, marked in black, form ADD$^+$-sites.

Asp f 2: BALB/c, CBA, C57BL/6, human IgE

DAGAVTSFPIHSSCNATQ**RR**QIEAGLNEAVELA**R**HAKAHILRWGNESEIY
RKYFGN**R**PTMEAVGAYDVIVNGDKANVLFRCDNPDGNCALEGWGGHWRGA
NATSETVICDRSYTT**RR**WLVSMCSQGYTVAGSETNTFWASDLMHRLYHVP
AVGQGWVDHFADGYDEVIALAKSNGTESTHDSEAFEYFALEAYAFDIAAP
GVGCAGESHGPDQGHDTGSASAPASTSTSSSSSGSGSGATTTPTDSPSAT
IDVPSNCHTHEGGQLHCT

Asp f 3: BALB/c, C57BL/6, human IgE

MSGLKAGDSFPSDVVFSYIPWSEDKGEITACGIPINYNASKEWAD**KK**VIL
FALPGAFTPVCSARHVPEYIEKLPEI**RAK**GVDVVAVLAYNDAYVMSAWG**K**
ANQVTGDDILFLSDPDARFSKSIGWADEEGRTKRYALVIDHG**K**ITYAALE
PAKNHLEFSSAETVLKHL

Fig. 2. Overlay of B-cell epitopes and IDIC-diagrams for Asp f 2 and Asp f 3. ADD^+-sites are shaded. Positively charged aa R and K within or in close proximity to ADD^+-sites, recognized by Ig, are shown in bold. Linear B-cell epitopes in Asp f 2 and Asp f 3 recognized by IgG from BALB/c, CBA and C57BL/6 mice and by IgE from ABPA patients are delineated. Dominant B-cell epitopes recognized by three strains of mice and human IgE are delineated with a thick line. More than one epitope can be located in delineated areas.

presumably outside of ADD^+-sites and 20% in ADD^+-sites containing positively charged amino acids arginine (R) or lysine (K). Among Asp f 3 B-cell epitopes only 52% were found outside of ADD^+-sites and the rest – in ADD^+-sites containing K/R.

None of the Asp f 2 or Asp f 3 B-cell epitopes were found within ADD^+-sites lacking R and/or L. Thus, we concluded that fragments of proteins residing in ADD^+-sites lacking R/K are less able to bind IgD expressed on the surface of B-cells and thus are not able to form B-cell epitopes.

3.3. *Selection of cryptic B-cell epitopes of Asp f 2, Der p 1, and Fel d 1*

The results described above have shown that it is likely that any peptide residing outside of ADD^+-sites lacking R/K can mount IgG response when delivered in an appropriate form. To study this hypothesis we have selected two peptides from Asp f 2 –Af-60–71 and Af-235–249, which are not natural B-cell epitopes (Table 2) and have expressed them in a recombinant form as VLPs. As a control Asp f 2 peptide – Af-254–268, which represents dominant B-cell epitope, was also obtained in a recombinant form. Besides four more peptides were selected from Der p 1 and Fel d 1.[18,19] These proteins are major allergens from house dust mites and cat dander, accordingly.

Peptides from major allergens are attractive candidates for anti-allergic vaccine production. Due to presumably low restriction on the selection of cryptic B-cell epitopes, peptides were designed to decrease potential autoimmune activity. As many major allergens are enzymes they have some homology with human antigens.

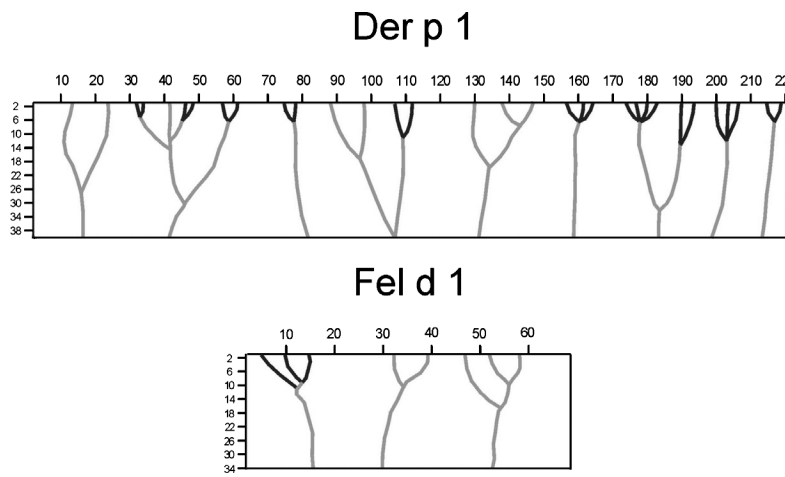

Fig. 3. IDIC-diagrams of Dep p 1 and Fel d 1. IDIC-branches, marked in black, form ADD$^+$-sites.

Asp f 2

DAG**A**VTSFPIH**S**SCNATQ**R**RQIE**A**G**L**NEAV**E**LARHAK**A**HILRWGNE**S**EIY
RKYFG**N**RPTMEA**V**GAYDVI**V**NG**D**KANVLFRCDNP**D**GNCAL**E**GWGGHWRGA
NAT**S**ETVICDR**S**Y**T**TRRWL**V**SMCSQGYTVA**G**SETNTFWA**S**DLMHRLYHVP
AVGQGWVDHFAD**G**YDEVIA**L**AK**S**NGTESTHDS**E**AFEYFA**L**EAYAFDIAA**P**
GVGCA**G**ESHG**P**DQGHDTGSA**S**APASTSTSS**S**SSG**S**GSGATTTPTDSP**S**AT
IDVPSNCHTHE**G**GQLHCT

Der p 1

TNACSINGNAPAEIDLRQMRTVTPIRMQGGCGSCWAFSGVAATESAYLAY
RNQSLDLAEQELVDCASQHGCHGDTIPRGIEYIQHNGVVQESYYRYVA**R**E
QSC**RR**PNAQ**R**FGISNYCQIYPPNVNKIREALAQTHSAIAVIIGI**K**DLDAF
RHYDG**R**TIIQ**R**DNGYQPNYHAVNIVGYSNAQGVDYWIVRNSWDTNWGDNG
YG**Y**FAANIDLMMIE**E**YPY**V**VIL

Fel d 1 (chain 1)

EICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAKMT
EEDKENALSVLDKIYTSPLC

Fig. 4. Overlay of selected peptides from Asp f 2, Der p 1 and Fel d 1 and protein information structure. ADD$^+$-sites are shaded. Peptides are delineated with a thick line. R/K in ADD$^+$ -sites are shown in bold.

Thus, to avoid potential autoimmune response all peptides were selected from parts non-homologous to human proteins. In the IS of proteins selected peptides were either completely outside of ADD$^+$-sites (Asp f 2 and Fel d 1) or included ADD$^+$-sites containing K/R (Der p 1). IDIC-diagrams of allergens and peptide sequences are shown in Figs. 3 and 4.

3.4. *Immunogenic properties of cryptic B-cell epitopes*

To study immunogenic properties of these peptides they were produced in a recombinant form as VLPs, chosen as both a vehicle and source of T-cell epitopes. Mice were immunized with VLP peptides and sera collected two weeks after the immunization and tested for the ability of allergen extracts to bind to the sera. The results indicate that all peptides produce a comparable IgG response (Fig. 5). There was no significant difference between the immune responses induced by peptides representing natural B-epitope (Asp f 2 254–268) and by peptides selected arbitrary outside of ADD^+-sites or including ADD^+-sites rich for R/K. Thus, the structural analysis of protein sequences can be used to identify ADD^+-sites, and especially

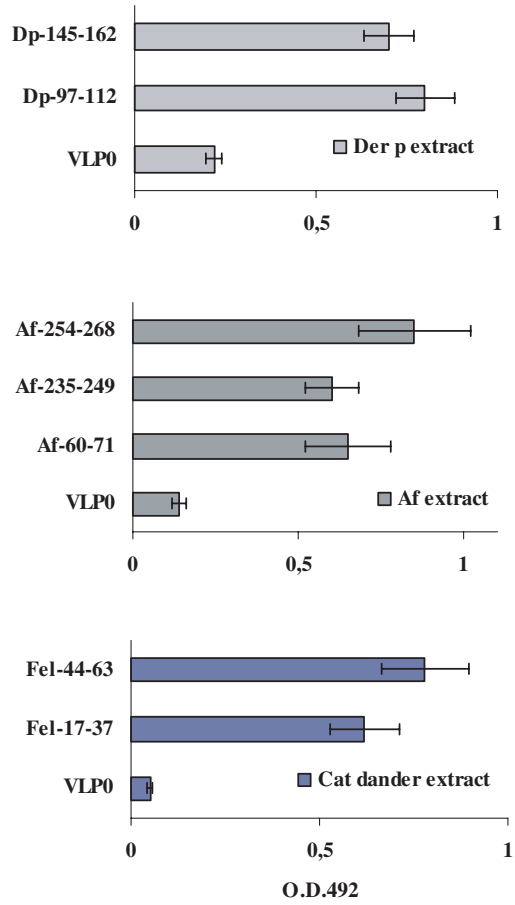

Fig. 5. Induction of IgG1 to Asp f 2, Der p 1 and Fel d 1 in B/c mice immunized with VLP-peptides. Peptide Af-254–268 represents a natural dominant B-cell epitope of Asp f 2. Other peptides are selected basing on IS structure of proteins and sequence homology to human antigens. Sera were diluted 1:500.

those, that do not contain R or K. It is likely that any peptides selected on the basis of IS analysis can mount humoral response when these peptides are delivered in combination with a T-cell epitope.

4. Discussion

Identification of antigenic regions (T- and B-cell epitopes) that are able to mount immune response and induce long lasting immunological memory is the most important and expensive stage of vaccine design. Up to now there are several data bases containing more than 1,500 known epitopes from viral, bacterial, fungal and protozoa antigens.[20,21] However, even the best set of prediction methods perform only marginally better than random selection.[5]

In this work a new method for cryptic B-cell epitope selection is suggested. It is based on the analysis of the information structure of protein sequences. IS analysis constructs hierarchical structures of proteins consisting of IDIC-sites, which form IDIC-trees. It was shown that IS of protein sequences has irregular structure with either increased or decreased density of IDIC-branches in the lowest hierarchic level (ADD^+- and ADD^--sites). It was hypothesized that these ADD-sites of proteins can have different immunogenic properties. This hypothesis was studied using 41 known B-cell epitopes from two proteins: Asp f 2 (24 epitopes) and Asp f 3 (17 epitopes). It was found that 97.3% of B-cell epitopes locate either outside of ADD^+-sites (67.6%) or in ADD^+-sites containing positively charged residues arginine and/or lysine (29.7%). The number of ADD^+-sites varies from protein to protein and covers on average 23% of protein sequences. The results obtained demonstrate that B-cell epitopes are scattered over the sequences without special requirements for binding for the exception of probably ADD^+-sites lacking R/K. No B-cell epitopes were found in these areas. On the contrary, among ADD^+-sites 54% lack R or K.

The analysis of B-cell epitope location has shown that peptides residing in ADD^+-sites are not immunogenic for the exception of those that contain positively charges residues arginine and/or lysine. As it is likely that for B-cell interaction with an antigen some level of protein chain flexibility is essential it can be hypothesized that this conformational rigidity is the reason why these segments of proteins locating in ADD^+-sites are not immunogenic. Availability of positively charged residues probably facilitates the binding of B-cells via electrostatic interactions.

Based on these results we have selected six peptides from Asp f 2 and allergens Der p 1 and Fel d 1. One Asp f 2 peptide represented natural dominant B-cell epitope while the others were selected arbitrarily on the basis of IS analysis as residing outside of ADD^+-sites lacking K/R. Recombinant peptides were constructed as yeast virus-like particles (VLPs) to provide a source of T-cell epitopes. Immunization of mice with all constructs has induced a comparable and significant humoral immune response. It is not clear why dominant B-cell epitopes are not more immunogenic than cryptic epitopes. It is likely that the natural dominance of B-cell epitopes is

more the result of T-cell help limitation when inability of a particular B-cell to recognize a cryptic epitope.

Thus, polypeptides residing outside of ADD^{+}-sites lacking R/K, which represent less then 10% of a protein sequence, can be considered as potential B-cell epitopes able to induce IgG, when injected in combination with T-cell epitopes. Low restriction of B-cell epitope selection is in accordance with the results obtained by Blythe & Flower,[5] where around 500 amino acid propensity scales were used to examine epitope location within 50 proteins mapped for polyclonal responses. It was concluded that single-scale amino acid propensity profiles (i.e. single requirement for B-cell to recognize any epitope) cannot be used to predict epitope location reliably.

Another support to the data obtained in this paper came from the work of Wu *et al.*,[22] where it was shown that 15-mer peptide derived from Der p 2, a major allergen of the house dust mite *D. pteronyssinus*, which contained a dominant T-cell epitope and was not recognized by antibodies to Der p 2, was able to elicit strong antibody response, suggesting the presence of a cryptic B-cell epitope.

Thus, a new method based on the analysis of information structure of protein sequences presented here can have a wide application for the vaccine design. However the validity of this approach and neutralizing potential of peptide vaccines produced on the basis of IS analysis should be carefully verified.

Acknowledgments

This publication was made possible in part by the grant 04-04-48815 from Russian FoundationforBasicResearch,RAS.

References

1. Hopp T, Woods K, Prediction of protein antigenic determinants from amino acid sequences, *Proc Natl Acad Sci USA* **78**:3824–3828, 1981.
2. Welling G, Weijer W, Zee R, Welling-Wester S, Prediction of sequential antigenic regions in proteins, *FEBS Lett* **188**:215–218, 1985.
3. Parker J, Guo D, Hodges R, New hydrophilicity scale derived from high-performance liquid chomatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites, *Biochemistry* **25**:5425–5432, 1986.
4. Novotny J, Handschumacher M, Haber E *et al.*, Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains), *Proc Natl Acad Sci USA* **83**:226–230, 1986.
5. Blythe MJ, Flower DR, Benchmarking B cell epitope prediction: underperformance of existing methods, *Protein Sci* **14**:246–248, 2005.
6. Rathore D, Nagarkatti R, Jani D *et al.*, An immunologically cryptic epitope of plasmodium falciparum circumsporozoite protein facilitates liver cell recognition and induces protective antibodies that block liver cell invasion, *J Biol Chem* **280**: 20524–20529, 2005.

7. Nekrasov AN, Analysis of information structure of protein sequences: A new method for analyzing the domain organization of proteins, *J Biomol Struct Dyn* **21**:615–623, 2004.
8. Rogov SI, Nekrasov AN, A numerical measure of amino acid residues similarity based on the analysis of their surroundings in natural protein sequences, *Protein Eng* **14**:459–463, 2001.
9. Nekrasov AN, Entropy of protein sequences: an integral approuch, *J Biomol Struct Dyn* **20**:87–92, 2002.
10. Bernard A, Coitot S, Bremont A, Bernard G, T and B cell cooperation: a dance of life and death, *Transplantation* **79**:S8–S11, 2005.
11. Benson EM, Clarkson J, Law M *et al.*, Therapeutic vaccination with p24-VLP and zidovudine augments HIV-specific cytotoxic T lymphocyte activity in asymptomatic HIV-infected individuals, *AIDS Res Hum Retroviruses* **15**:105–113, 1999.
12. Oliveira-Ferreira J, Miyahira Y, Layton GT *et al.*, Immunogenicity of Ty-VLP bearing a CD8(+) T cell epitope of the CS protein of *P. yoelii*: enhanced memory response by boosting with recombinant vaccinia virus, *Vaccine* **18**:1863–1869, 2000.
13. Svirshchevskaya E, Viskova N, Shevchenko M *et al.*, High-affinity IgG to a major A. fumigatus allergen, Asp f 2, retards allergic response, *Med Sci Monit* **10**:BR371–380, 2004.
14. Banerjee B, Kurup VP, Phadnis S *et al.*, Molecular cloning and expression of a recombinant Aspergillus fumigatus protein Asp f II with significant immunoglobulin E reactivity in allergic bronchopulmonary aspergillosis, *J Lab Clin Med* **127**:253–262, 1996.
15. Marchenko AN, Kozlov DG, Svirshchevskaya EV *et al.*, The p1 protein of the yeast transposon Ty1 can be used for constructing bifunctional virus-like particles, *J Mol Microbiol Biotechnol* **5**:97–104, 2003.
16. Banerjee B, Greenberger PA, Fink JN, Kurup VP, Conformational and linear B-cell epitopes of Asp f 2, a major allergen of Aspergillus fumigatus, bind differently to immunoglobulin E antibody in the sera of allergic bronchopulmonary aspergillosis patients, *Infect Immun* **67**:2284–2291, 1999.
17. Ramachandran H, Jayaraman V, Banerjee B *et al.*, IgE binding conformational epitopes of Asp f 3, a major allergen of Aspergillus fumigatus, *Clin Immunol* **103**: 324–333, 2002.
18. Takai T, Kato T, Yasueda H *et al.*, Analysis of the structure and allergenicity of recombinant pro- and mature Der p 1 and Der f 1: major conformational IgE epitopes blocked by prodomains, *J Allergy Clin Immunol* **115**:555–563, 2005.
19. Spitzauer S, Pandjaitan B, Muhl S *et al.*, Major cat and dog allergens share IgE epitopes, *J Allergy Clin Immunol* **99**:100–106, 1997.
20. Saha S, Bhasin M, Raghava GP, Bcipep: a database of B-cell epitopes, *BMC Genomics* **6**:79–85, 2005.
21. Schonbach C, Koh JL, Flower DR, Brusic V, An Update on the Functional Molecular Immunology (FIMM) Database, *Appl Bioinformatics* **4**:25–31, 2005.
22. Wu B, Toussaint G, Vander Elst L *et al.*, Major T cell epitope-containing peptides can elicit strong antibody responses, *Eur J Immunol* **30**:291–299, 2000.

Elena Svirshchevskaya is a Senior Research Scientist in Department of Immunology at Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences.

Her current research work is devoted to immunological disorders such as allergy and autoimmune diseases. Recently she has started a new project on the ability of different ways (immunomodulation, chemical treatment) to suppress tumour stem cells to treat breast cancer in mouse models.

Ludmila Alekseeva graduated from Moscow Institute of Fine Chemical Technology, Department of Organic Synthesis and Biotechnology in 1991 and joined the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, where at the present she is employed as a Scientific Researcher. In 2002 she finished her Ph.D. project on peptides which might be used for specific immunotherapy of allergy caused by fungus *Aspergillus fumigatus*. Her main scientific interests may be briefly defined as the study of different aspects of immune response development with peptides.

Alexei Marchenko has graduated from Moscow State University at 1997. He has obtained his Ph.D. in genetics in 2003. Since 1998 he has worked in State Research Institute for Genetics and Selection of Industrial Microorganisms on a wide variety of biotechnological topics. These have included a detailed investigation of virus-like particles assembling in *Saccharomyces cerevisiae* and the problem of pentose utilization by industrial strains *Escherichia coli*. His current work is on the creation of fungi strains — human hormones producers.

Sergei Benevolensky obtained a BSc in Microbiology in 1974 from Moscow State University, Russia and a Ph.D. in Genetics in 1980 in State Institute of Genetics and Selection of Industrial Microorganisms (SIG), Russia. Since 1980 he has worked in SIG in field of yeast genetics and biotechnology. He is currently working in SIG participating in several projects related to construction of yeast and fungi protein producing strains.

Valentina M. Berzhec, Professor, is a Leader of Laboratory of Allergens Development at Ilya Mechnikov Institute for Vaccine and Sera, Russia. Her current scientific interests are connected to development of new preparations of allergens for diagnostic and treatment of allergic diseases of different etiology.

Alexei Nekrasov is a Senior Research Scientist in Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences. His current research interests include computer modelling and three-dimensional structures of proteins and peptides.