Prediction of Epitopes in Closely Related Proteins Using a New Algorithm

I. I. Davydov, S. Fidalgo*, S. A. Khaustova, V. G. Lelyanova*, E. S. Grebenyuk, Yu. A. Ushkaryov*, and A. G. Tonevitsky

Translated from *Byulleten' Eksperimental'noi Biologii i Meditsiny*, Vol. 148, No. 12, pp. 628-632, December, 2009
Original article submitted March 18, 2009

Latrophilin 1 (presynaptic receptor) binds α-latrotoxin from black widow spider venom and regulates neurotransmitter release from nerve endings. The study of the mechanism of action of this receptor is impeded by the existence of closely related latrophilins 2 and 3. A profile of differences detecting the most differing and identical sites in several proteins was developed in order to obtain highly specific antibodies for differentiation between isoforms of related proteins. In addition, we used an algorithm for prediction of immunogenic sites of the protein, based on the basic vector method. The peptides selected using this algorithm were used for immunization of animals. The resultant sera exhibited the estimated specificity and high affinity for the corresponding receptor forms.

Key Words: *latrophilin; epitope; closely related proteins; basic vector method; amino acid pairs*

α-Latrotoxin (α-LTX) from black widow spider (*Latrodectus*) venom affects presynaptic neuronal endings and causes massive release of neurotransmitters in them. Due to its high activity, specificity, and capacity to induce exocytosis in the absence of Ca ions, α-LTX is now widely used as a tool for studies of the molecular mechanisms regulating presynaptic processes [9].

High-affinity interactions between α-LTX and presynaptic protein latrophilin 1 (also known as calcium-independent receptor for α-latrotoxin; CIRL) in the brain led to detection [2,3] and later characterization of this important receptor. It was shown, among other things, that stimulation of latrophilin 1 triggers intracellular signals leading to the release of calcium stores. Hence, latrophilin 1 regulates secretion of neurotransmitters [9]. However, the study of the role of this receptor is impeded by the existence of three closely related isoforms (latrophilins 1, 2, and 3) with different functions and by the absence of highly specific antibodies [9].

Antibodies are a potent tool for studies of proteins due to their capacity to high-affinity binding of the antigen (*e.g.*, protein). The protein site recognized by the antibody is called an epitope. Linear epitopes are formed by amino acids, forming a sequence, while conformation epitopes consist of amino acids located in different sites of the protein chain. The majority of antibodies to proteins bind conformational epitopes [7].

However, the whole protein is not an obligatory condition for obtaining polyclonal antibodies: just a small protein fragment is sufficient for this. Immunization with a short peptide results in production of antibodies to the protein with a preset epitope. However, it was shown that antibodies are not produced to any protein site. Therefore, prediction of protein sites to which antibodies can be produced is an important problem.

By the present time, several machine-learning approaches to the solution of this problem are available. One of the most effective approaches is the method

Address for correspondence: All-Russian Research Institute of Physical Culture and Sports Education, Moscow, Russia; *Division of Cell and Molecular Biology, Imperial College London, UK. davydov@bioinf.ru. I. I. Davydov

0007-4888/09/1486-0869

based on the incidence of amino acid pairs and the basic vectors method [1].

The aim of our study was to obtain antibodies, specifically recognizing the N-terminal domains of closely related rat latrophilins 1, 2, and 3, and antibodies binding the N-terminal domain of each latrophilin. This problem was solved by using the algorithm for selection of the optimal peptide epitopes in the amino acid sequences of latrophilins and subsequent immunization of rabbits with conjugates of the respective peptides and analysis of immune sera.

MATERIALS AND METHODS

The epitopes were predicted using the algorithm based on the basic vectors method and the incidence of amino acid pairs. The data from the BCIPEP database [8] supplemented by our experimental findings, served as the positive data set. Random peptides from the Swiss-Prot database [2] served as the negative set.

The incidence of amino acid pairs was calculated by the formula:

$$R_{AAP}=\log\left(\frac{f^{-}_{AAP}}{f^{+}_{AAP}}\right).$$

The scale values were then standardized for the [-1, +1] range:

$$R_{AAP}=2\left(\frac{R_{AAP}\text{-min}}{\text{max-min}}\right)\text{-1 [3].}$$

The incidence of amino acid pairs was inputted in the support vector machine with Gaussian type radial-basis function kernel:

$$K(x,x')=\exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right)$$

Parameters σ^2 and C (penalty) were selected experimentally; $\sigma^2=2$ and $C=32$ proved to be optimal. Realization of this algorithm is available (http://www.bioinf.ru/aappred/).

The similarity coefficients were obtained using the ClustalX software [5]. Paired one-by-one amino acid similarity tables for proteins were plotted and standardized in the [-3, +3] range. The coefficient of difference for each amino acid was determined as $3=\max(S_a, S_b)$, where S_a and S_b are similarity coefficients for the amino acid in a certain protein and the corresponding amino acid in each of the two other proteins. The resultant coefficients were averaged for the entire protein length using 10-amino acid averaging window for obtaining the difference profile.

Hence, the difference profile for each protein characterizes the degree of the maximum difference of this protein from two other ones (or, generally speaking, from all other proteins).

In addition to the individual difference profiles, the general profile of differences was calculated as the maximum *vs.* all available individual difference profiles.

The general difference profile characterizes general differences between the proteins. Protein sites where the general difference profile has minimum (-3) values are identical.

Peptides were conjugated with KLH protein via MBS. Rabbits were immunized with repeated (at 2-week interval) injections of 1 ml respective conjugate in a concentration of 0.5 mg/ml. The sera (20 ml) were collected on days 7 and 10 after each immunization. Specific antibodies were isolated from the sera by biospecific chromatography on adsorbents conjugated with the corresponding peptide.

The secreted fragment of latrophilin 1 corresponding to the entire extracellular N-terminal domain (849 amino acids, including artificially inserted epitopes [10]) was expressed in steadily transfected neuroblastoma cells cultured in serum-free medium for the latest 48 h. Aliquots of the medium containing 10-20 ng latrophilin 1 were fractionated in 8% PAAG in parallel with specimens of rat forebrain and cerebellum dissolved in 1% Triton and brought to the same protein concentration (20 μg/sample). Protein bands were transferred from the gel onto PVDF membrane. Identical blots thus obtained were incubated separately with each primary antibody (2 μg/ml) and then with second goat antibodies to rabbit immunoglobulin, conjugated with horseradish peroxidase. Positive fluorescent signals were developed by the standard method in accordance with the instruction (Millipore).

In order to detect latrophilin on the cell surface, full-length latrophilin 1 was expressed in neuroblastoma cells cultured on coverslips. The cells were fixed and incubated with peptide antigen (20 μg/ml), then fixed, permeabilized, and incubated with mouse monoclonal antibodies to myc epitope located on the intracellular C-terminal domain of latrophilin 1 [10]. The cells were then incubated simultaneously with two second antibodies: to rabbit immunoglobulin conjugated with Alexa 488 fluorophore (green) and to mouse immunoglobulin conjugated with Alexa 594 fluorophore (red); the coverslips with the cells were then fixed on slides. The images were recorded using LSM510 laser scanning adapter connected to Axioplan 2 microscope (Zeiss).

RESULTS

For prediction of epitopes, the difference profiles for N-terminal domains of each latrophilin and the general profile of differences for N-terminal domain of

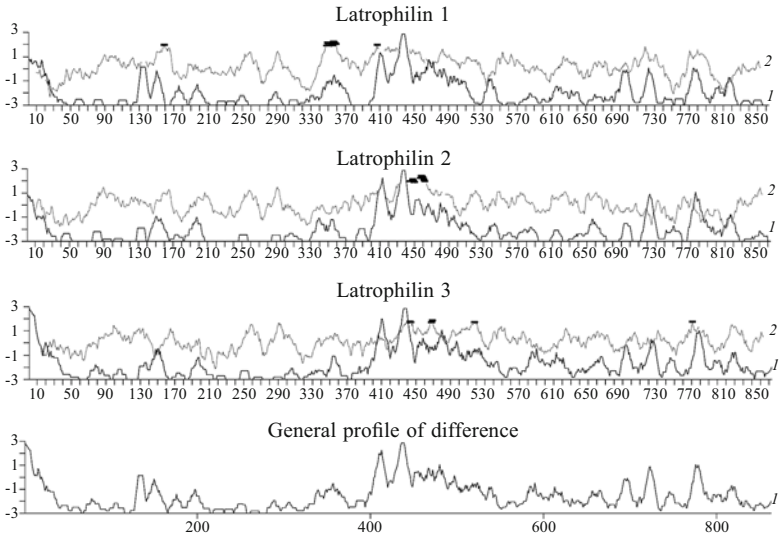

Fig. 1. Profile of differences and profile of antigenic activity. *1*) profiles of differences for N-terminal domains of latrophilin 1, latrophilin 2, latrophilin 3, and general profile of difference. *2*) profile of antigenicity of protein. Black dash: sites of antigenicity profile surpassing the threshold value.

all three proteins were plotted (Fig. 1). The value of 1.9 was taken for the threshold antigenicity value. Since the maximum antigenicity profile of N-terminal domain OF latrophilin 3 was below 1.9, the value of 1.7 was chosen for its threshold.

The detected sites with high antigenicity were ranked by the values of difference profile (Table 1). Peptides 407 and 443-444 were excluded because of their possible O-glycosylation. Peptide 445-462 was shifted by three amino acids towards the N-terminal so that the resultant antibodies could recognize also bovine latrophilin 2. Peptide 260-275 was selected for obtaining antibodies recognizing latrophilins 1, 2, and 3 because of the antigenic activity of the site with the lowest general profile of difference.

Based on these data, the following peptides were synthesized for immunization of animals: CYAFNTNANREEPVSLAFPNP (latrophilin 1); CQRGPVSSTVAGPQEGSRGTK (latrophilin 2); CSTTPSLPGRRNRSTSTPSPA (latrophilin 3); and CHDTSPYRWGGKTDIDLAVDE (latrophilins 1, 2, and 3). The corresponding sera were called PAL1, PAL2, PAL3, and PAL123. The sera to all peptides obtained after immunization 5 had 1:30,000 titer and specifically interacted in EIA with the corresponding peptide conjugated with BSA.

TABLE 1. Immunogenic Sites of Protein

No.	Antigenic activity	Coefficient of difference	Sequence
PAL1			
407	1.96	-0.03	GPPDPSAGPATSPPLSTTT
355-359	2.2	-0.51	VDYAFNTNANREEPVSLAFPNPYQF
348-351	2.14	-1.01	SEAAGNRVDYAFNTNANREEPVSL
159	1.95	-2.51	LEPTSTHESEHQSGAWCKDP
PAL2			
445-462	2.41	0.5	AELFKTTVSTTSSTSQRGPVSSTVAGPQEGSR
PAL3			
443-444	1.77	0.7	STTGPLGMGSTTTSTTLRTTT
468-469	1.87	-0.33	GRSTTPSLPGRRNRSTSTPSPA
518	1.72	-0.51	AREIMWFKTRQGQVAKQSCPA
771	1.76	-1.34	YLSTENASMKLGTEAMSTNHSV
PAL123			
260-275	0.69	-3	HDTSPYRWGGKTDIDLAVDENGLWV

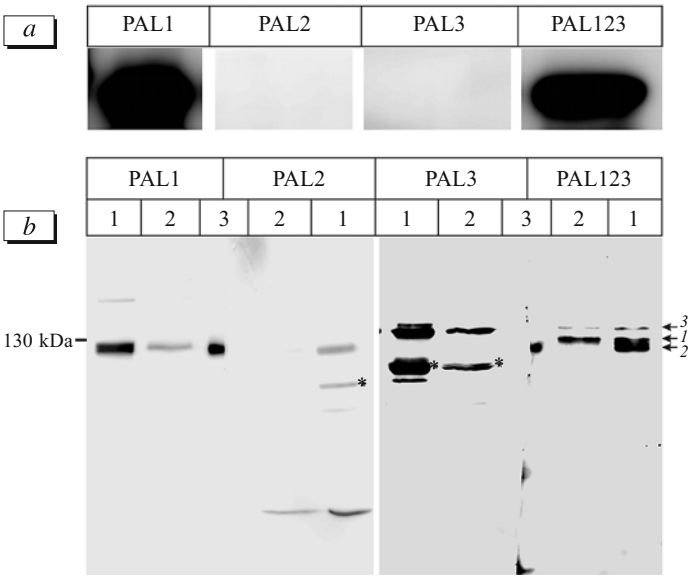

Fig. 2. Immune staining of latrophilins with PAL antibodies. *a*) Immune staining of latrophilin 1 recombinant N-terminal domain with PAL1, PAL2, PAL3, and PAL123 antibodies. Antibody concentration 20 μg/ml. The duration of incubation was selected to detect even the slightest cross-reaction. *b*) latrophilin immunoblotting in extracts of the rat brain. *1*) forebrain extract; *2*) cerebellum extract; *3*) latrophilin 1 recombinant N-terminal fragment (control). Peptide antibodies were used in concentrations: PAL1: 2 μg/ml; PAL2, PAL3, and PAL123: 4 μg/ml. Figures on the right (*1*, *2*, *3*) indicate protein bands corresponding to latrophilins 1, 2, and 3. Asterisks show splice-variants of latrophilin.

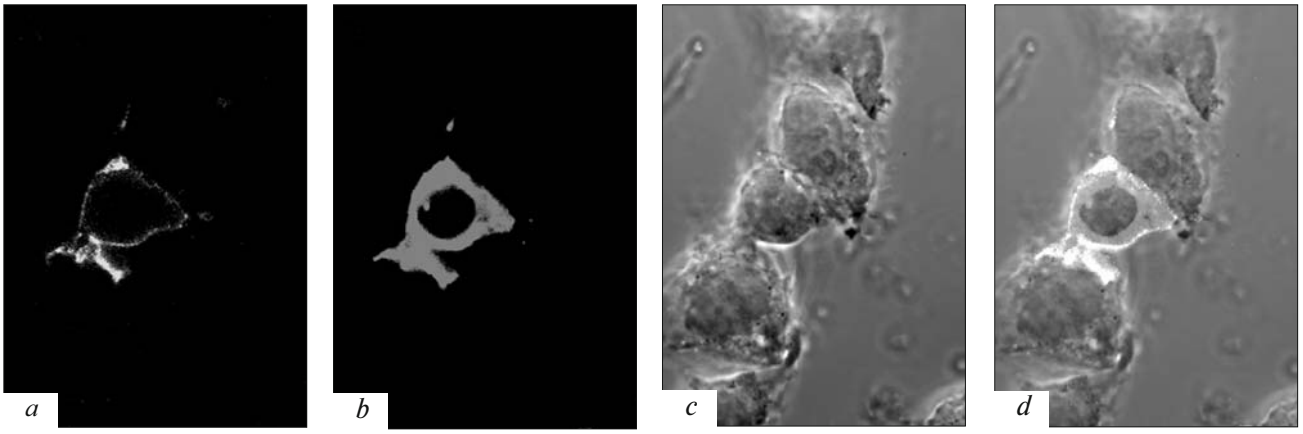

Fig. 3. Detection of latrophilin 1 expressed on cell surface. Indirect immunodetection with PAL1 antibodies and monoclonal antibodies to myc epitope on latrophilin C-terminal domain. a) second antibodies to PAL1; *b*) second antibodies to monoclonal antibodies; *c*) cells; *d*) cells and both antibody types.

Immune staining of cells (Fig. 2) showed specificity of recognition of N-terminal domain latrophilin 1 expressed in cell culture and isolated from rat brain by PAL1 and PAL123 antibodies. Affinity of PAL1 antibodies to latrophilin 1 was higher than that of PAL123, which agrees with predicted antigenicities of the peptides (Table 1). Antibodies PAL2 and PAL3 even in high concentrations gave no nonspecific reaction with latrophilin 1 (Fig. 2, *a*).

Antibodies PAL1 demonstrated the capacity to stain native latrophilin 1 on the cell surface (Fig. 3). However, affinity of this interaction was lower than in the reaction with denaturated protein. Presumably, this was due to the fact that the site was not completely exposed on the surface of native protein.

These results prove that the algorithm of epitope prediction on the basis of machine learning approach used in our study effectively detects the most immunogenic sites of protein. This algorithm in combination with the difference profile makes it possible to obtain specific antibodies even for closely related proteins.

We are grateful to Dr. I. V. Shpilevaya and Dr. A. A. Dokrunova for assistance in antibody obtaining and purification.

The study was supported by the British Royal Society (International Collaborative Project/R4) and Wellcome Trust Charity Organization (GR083199).

REFERENCES

1. I. I. Davydov and A. G. Tonevitsky, *Mol. Biol.*, **43**, No. 1, 166-174 (2009).
2. A. Bairoch, B. Boeckmann, S. Ferro, and E. Gasteiger, *Brief. Bioinform.*, **5**, No. 1, 39-55 (2004).
3. J. Chen, H. Liu, J. Yang, and K. C. Chou, *Amino Acids*, **33**, No. 3, 423-428 (2007).

4. V. G. Krasnoperov, M. A. Bittner, R. Beavis, *et al.*, *Neuron*, **18**, No. 6, 925-937 (1997).
5. M. A. Larkin, G. Blackshields, N. P. Brown, *et al.*, *Bioinformatics*, **23**, No. 21, 2947-2948 (2007).
6. V. G. Lelianova, B. A. Davletov, A. Sterling, *et al.*, *J. Biol. Chem.*, **272**, No. 34, 21,504-21,508 (1997).
7. J. Pellequer, E. Westhof, and M. van Regenmortel, *Methods Enzymol.*, **203**, 176-201 (1991).
8. S. Saha, M. Bhasin, and G. P. Raghava, *BMC Genomics*, **6**, No. 1, 79 (2005).
9. Y. A. Ushkaryov, A. Rohou, and S. Sugita, *Handb. Exp. Pharmacol.*, No. 184, 171-206 (2008).
10. K. E. Volynski, J. P. Silva, V. G. Lelianova, *EMBO J.*, **23**, No. 22, 4423-4433 (2004).