Discovery of New Therapeutic Targets by the Informational Spectrum Method

Nevena Veljkovic[1], Sanja Glisic[1], Jelena Prljic[1], Vladimir Perovic[1], Maurizio Botta[2] and Veljko Veljkovic[1,*]

[1]*Center for Multidisciplinary Research, Institute of Nuclear Sciences VINCA,P.O.Box 522, 11001 Belgrade, Serbia;* [2]*Dipartimento Farmaco-Chimico Tecnologico, Università degli Studi di Siena, Via A. Moro, 53100 Siena, Italy*

Abstract: The field of bioinformatics has become a major part of the drug discovery pipeline playing a key role in improvement and acceleration of this time and money consuming process. Here we review the application of the informational spectrum method (ISM), a virtual spectroscopy method for structure/function analysis of proteins, in identification of functional protein domains representing candidate therapeutic targets for drugs against human immunodeficiency virus (HIV)-1, anthrax, highly pathogenic influenza virus H5N1 and cardiovascular diseases.

Keywords: Bioinformatics, target, electron-ion interaction potential, informational spectrum method, H5N1, HIV-1, anthrax.

INTRODUCTION

Target discovery, which involves the identification and early validation of therapeutic targets, is an essential first step in the drug discovery process which on average takes twelve to fifteen years to complete, and costs approximately 800 million dollars. Understanding how biomolecules recognize each other is very important for identification of protein and nucleic acid targets and prediction of drug-protein and drug-nucleic acid interactions, because this knowledge can significantly facilitate the lengthy and expensive process of drug discovery.

The surface complementarity between interacting biomolecules, which was originally proposed by Emil Fischer in 1894, together with the collision theory, assuming that the first contact between interacting molecules is achieved accidentally by the thermal motions that cause molecular wander, represents the fundamental basis for our current understanding of intermolecular interaction in biological systems. If proteins are considered as spheres of 18 Å radius (typical of a small protein), and if spheres associate with every contact, without regard to orientation, the diffusion-limited association rate constant, calculated according to Smoluchowski's equation [1] is 7 x 10^9 $M^{-1}s^{-1}$. However, before chemical bond formation takes place, reacting molecular regions must be positioned close enough (at a distance of ≈ 2Å) and the appropriate atoms must be held in the correct orientation for the reaction that is to follow, because the attractive forces involved in the recognition and binding of molecules include all the weak non-covalent forces. It means that the protein's binding site is only a small fraction (~0.1%) of the surface area. Taking into account this limitation, the diffusion-limited association rate constant, predicted from a three-dimensional (3D) "random diffusion" model and calculated according Smoluchowski's equation is ~ 10^6 $M^{-1}s^{-1}$ for a protein-ligand and ~ 10^3 $M^{-1}s^{-1}$ for a protein-protein interaction. Northrup and Erickson have noted that protein-protein association generally occurs at rates that are 10^3 to 10^4 times faster than would be expected from simple considerations of collision frequencies and strict orientation effects which assume that productive binding occurs only when the molecules collide within 2 Å of their final binding site [2]. Taking into account many protein associations, including the very general reactions of antibody with protein antigen, these authors proposed ~10^6 $M^{-1}s^{-1}$ as a typical real rate for proteins associating and docking at the precise orientation for bond formation [2]. Some reactions are much faster than would be predicted from a 3D "random diffusion" model. A classic example is superoxide dismutase which has a catalytic rate constant of ~ 10^9 $M^{-1}s^{-1}$ even though the active site is only ~0.1% of the surface area [3].

In order to overcome the discrepancy between theoretically estimated values and real values of the associated rate constant for a protein-protein interaction, several models have been proposed (e.g. reducing the dimensionality of diffusion [4-6], protein association based on hydrodynamic steering [7], desolvation-mediated protein-protein binding [8], etc.). One of the most acceptable solutions for this problem was proposed by the physicist Fröhlich [9-11]. Starting from the very general theoretical considerations, he proposed that biological macromolecules should be capable to excite dipole vibrations. The interactions between biological molecules in a polar medium, including water and lipids, based on such vibrations lead to the appearance of frequency-selective long-range attractive forces which are efficient at a distance longer than one linear dimension of the interacting macromolecules. The consequence of such frequency complementarity between interacting molecules is an increase of numbers of productive collisions in comparison with the accidental encounter of molecules.

*Address correspondence to this author at the Center for Multidisciplinary Research, Institute of Nuclear Sciences VINCA, P.O.Box 522, 11001 Belgrade, Serbia; Tel: +381 11 24 53 686; Fax: + 381 11 34 40 100;
E-mail: vv@vin.bg.ac.yu

1389-2037/08 $55.00+.00

Accordingly, each of these properties, the long-range frequency compatibility and the structural complementarity, represents a necessary but not a sufficient condition for efficient biochemical reaction. Only both properties combined enable the efficient interaction between biological molecules. What does this mean from a practical point of view? The drug which perfectly fits the structure of the target molecule but without the "address" information which allows it long-range recognition of this target, cannot be efficient. It is like fishing with a bare fish-hook. Without an appropriate decoy, fishing cannot be successful, irrespective of how the fish-hook is designed.

AQVN/EIIP CONCEPT

It has been proposed that the number of valence electrons and the electron-ion interaction potential (EIIP), representing the main energy term of the valence electrons, are essential physical parameters of biological molecules determining their long-range properties [12]. We showed [12,13] that EIIP can be determined for organic molecules by the following simple equation derived from the "general model pseudopotential" [14-16]:

$$W = 0.25Z^*\sin(1.04\pi Z^*)/2\pi \quad (1)$$

where Z* is the average quasivalence number (AQVN) determined by

$$Z^* = \Sigma n_i Z_i / N \quad (2)$$

where Z_i is the valence number of the i-th atomic component, n_i is the number of atoms of the i-th component, m is the number of atomic components in the molecule, and N is the total number of atoms. The EIIP values calculated according to equations (1) and (2) are in Rydbergs (Ry). A strong connection has been demonstrated between EIIP and AQVN of organic molecules and their biological activity (mutagenicity, carcinogenicity, toxicity, antibiotic and cytostatic activity, etc.) [12,13,17-22].

THE INFORMATIONAL SPECTRUM METHOD

The long-range bio-molecular interactions represent an important factor which influences biological processes. For this reason, long-range interactions representing intrinsic physicochemical properties of proteins and nucleotide sequences should be included in analysis of protein-protein and protein-DNA interactions. As we described above, EIIP represents an important physical parameter which determines long-range properties of biological molecules. The EIIP is used for the development of the informational spectrum method (ISM), a virtual spectroscopy method for investigation of protein-protein and protein-DNA interactions, as well as for analysis of structure/function relationship of proteins and nucleotide sequences. This virtual spectroscopy method involves three basic steps exemplified by analyzing the interaction between the tumor suppressor p53 and ubiquitin-protein ligase Mdm2 (Fig. (**1**)):

1. Transformation of the alphabetic code of the primary structure into a sequence of numbers by assigning to each amino acid or nucleotide a corresponding numerical value representing the electron-ion interaction potential.
2. Conversion of the obtained numerical sequence by Fourier transformation into the informational spectrum.
3. Cross-spectral analysis which allows identification of frequency components in the informational spectrum of molecules which are important for their biological function or interaction with other molecules.

The physical and mathematical basis of ISM was described in details elsewhere [23-26], and here we will only present this bioinformatics method in brief. A sequence of N residues is represented as a linear array of N terms, with each term given a weight. The weight assigned to a residue is EIIP (Table **1**). In this way the alphabetic code is transformed into a sequence of numbers. The obtained numerical sequence, representing the primary structure of protein, is then subjected to a discrete Fourier transformation, which is defined as follows:

$$X(n) = \Sigma x(m)e^{-j(2/N)nm},\ n = 1, 2, ..., N/2 \quad (3)$$

where x(m) is the m-th member of a given numerical series, N is the total number of points in this series, and X(n) are discrete Fourier transformation coefficients. These coefficients describe the amplitude, phase and frequency of sinusoids, which comprised the original signal. The absolute value of complex discrete Fourier transformation defines the amplitude spectrum and the phase spectrum. The complete information about the original sequence is contained in both spectral functions. However, in the case of protein analysis, relevant information is presented in an energy density spectrum [23,24], which is defined as follows:

$$S(n) = X(n)X^*(n) = |X(n)|^2,\ n = 1, 2, ..., N/2. \quad (4)$$

In this way, sequences are analyzed as discrete signals. It is assumed that their points are equidistant with the distance d = 1. The maximal frequency in a spectrum defined in this way is F = 1/2d = 0.5. The frequency range is independent of the total number of points in the sequence. The total number of points in a sequence influences only resolution of the spectrum. The resolution of the N-point sequence is 1/n. The n-th point in the spectral function corresponds to a frequency f(n) = nf = n/N. Thus, the initial information defined by the sequence of amino acids can now be presented in the form of an informational spectrum (IS), representing a series of frequencies and their amplitudes.

The IS frequencies correspond to distribution of structural motifs with defined physicochemical properties determining a biological function of a protein. When comparing proteins, which share the same biological or biochemical function, the ISM technique allows detection of code/ frequency pairs which are specific for their common biological properties, or which correlate with their specific interaction. This common informational characteristic of sequences is determined by a cross-spectrum or consensus informational spectrum (CIS). A CIS of N spectra is obtained by the following equation:

$$C(j) = \Pi S(i,j) \quad (5)$$

where Π(i,j) is the j-th element of the i-th power spectrum and C(j) is the j-th element of CIS. Thus, CIS is the Fourier transform of the correlation function for the spectrum. In this way, any spectral component (frequency) not present in all compared informational spectra is eliminated. Peak frequen-

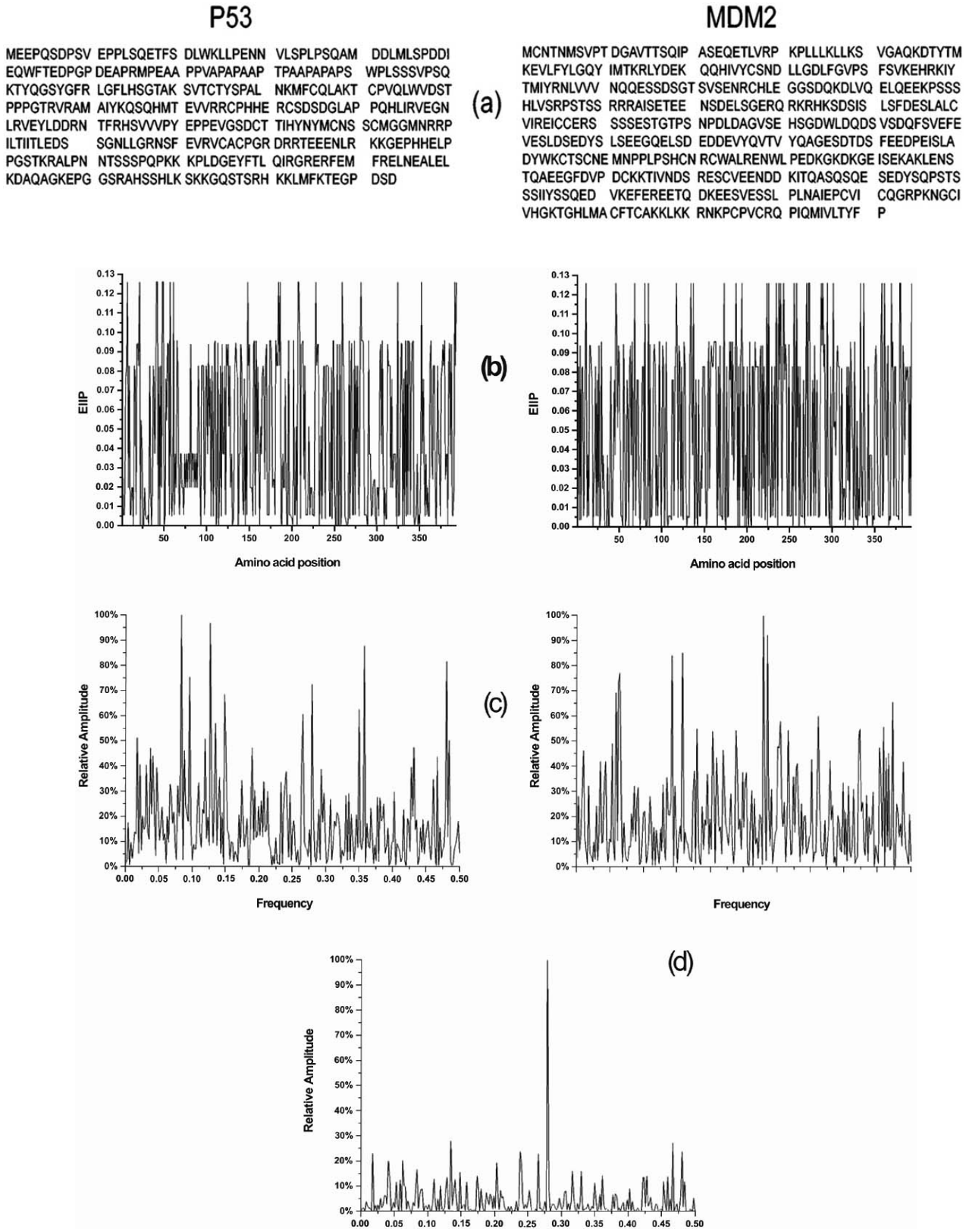

Fig. (1). Informational spectrum analyses of p53 and MDM2 proteins. (a) The alphabetic presentation of primary structures of of p53 and MDM2 molecules. **(b)** The numerical presentation of primary structures of of p53 and MDM2 molecules. **(c)** The spectral presentation of primary structures of of p53 and MDM2 molecules. **(d)** cross-spectrum of p53 and MDM2 molecules. The abscissa represents the frequencies from the Fourier transform of the sequence of EIIP. The lowest frequency is 0.0 and the highest is 0.5. The ordinate represents amplitudes, in arbitrary units, corresponding to each frequency component in the informational spectrum (IS), normalized on the maximal amplitude value.

Table 1. EIIP Values of Amino Acids and Nucleotides

Amino acid	EIIP [Ry]
Leu	0.0000
Ile	0.0000
Asn	0.0036
Gly	0.0050
Glu	0.0058
Val	0.0057
Pro	0.0198
His	0.0242
Lys	0.0371
Ala	0.0373
Tyr	0.0516
Trp	0.0548
Gln	0.0761
Met	0.0823
Ser	0.0829
Cis	0.0829
Thr	0.0941
Phe	0.0946
Arg	0.0959
Asp	0.1263
Nucleic acid	
A	0.1260
T	0.1335
G	0.0806
C	0.1340

cies in CIS are common frequency components for the analyzed sequences. A measure of similarity for each peak is a signal-to-noise ratio (S/N), which represents a ratio between signal intensity at one particular IS frequency and the main value of the whole spectrum. If one calculates a CIS for a group of proteins, which have different primary structures, and finds strictly defined peak frequencies, it means that the analyzed proteins participate in mutual interaction or have a common biological function.

The ISM was successfully applied in structure-function analysis of different protein and DNA sequences, as well as in *de novo* design of biologically active peptides [27-77].

APPLICATION OF ISM IN THE DISCOVERY OF NEW THERAPEUTIC TARGETS

Selection of Peptide HIV-1 Entry Inhibitors [74]

In spite of high variability the HIV-1 envelope glycoprotein 120 (gp120) preserves its tropism for the CD4 receptor and principal coreceptors CCR5 and CXCR4. It means that some property of gp120 which is responsible for its interaction with these receptors is highly conserved. In order to identify this conserved property which determines specific interaction between HIV-1 and receptor CCR5, the gp120 molecules from 50 R5 HIV-1 isolates have been subjected to ISM analysis. The obtained CIS, which is presented in Fig. (**2a**), contains only one characteristic peak corresponding to the IS frequency F(0.1035). This peak frequency represents common biological information encoded by the primary structure of analyzed gp120 molecules. According to the ISM concept this frequency probably represents information which is responsible for interaction between gp120 and the CCR5 receptor. In order to prove this assumption, we multiplied the consensus informational spectrum of gp120 with the informational spectra of CCR5 and CXCR4 receptors. As can be seen from results of this cross-spectral analysis (Fig. (**2b**) and Fig. (**2c**)), the amplitude corresponding to the frequency F(0.1035) significantly increases after multiplication of CIS gp120 with IS of CCR5 and decreases after its multiplication with IS of the CXCR4 receptor. This result confirms that only CCR5 shares common information with the consensus informational spectrum of analyzed gp120 molecules, and points out frequency F(0.1035) as important for gp120/CCR5 interaction.

Further computer scanning survey of the CCR5 primary structure revealed that the N-terminus of the second extracellular loop (ECL2) encompassing residues 168 to 186 is essential for information represented by the IS frequency F(0.1035). This result was further used for selection of the peptide HIV entry inhibitors. For this purpose, phage display screening was performed using as a target the CCR5-derived peptide encompassing residues 168 to 186. In this way a group of peptides, seven amino acids long, has been selected as candidate HIV-1 entry inhibitors which block gp120/CCR5 interaction. In Fig. (**3**) the results of HIV-1 entry inhibition with the most active candidate peptide inhibitor (denoted as peptide P25) are presented [74]. In a nanomolar range this peptide was even more effective than peptide T-20, representing the first commercially available HIV-1 entry inhibitor.

Identification of Human Proteins that Bind the Anthrax Toxin [75]

Protein capillary morphogenesis protein 2 (CMG2) was reported as the anthrax receptor [78]. Cross-spectrum between CMG2 and the protective antigen of bacillus anthraces, which is responsible for delivery of anthrax toxin into the cell, is given in Fig (**4a**) [75]. The dominant peak in this cross-spectrum corresponds to F(0.087) pointing out this IS frequency as important for interaction between the CMG2 receptor and the anthrax toxin. The cross-spectral analysis between the protective antigen and over 13,000 human proteins from the UniProt database [79] revealed elastin microfibril interfase located protein (EMILIN)-1 as a human protein which in the cross-spectrum with the protective antigen has the highest amplitude at frequency F(0.087) (Fig. (**4b**)). This result singled out EMILIN-1 as a candidate binding protein for the anthrax toxin.

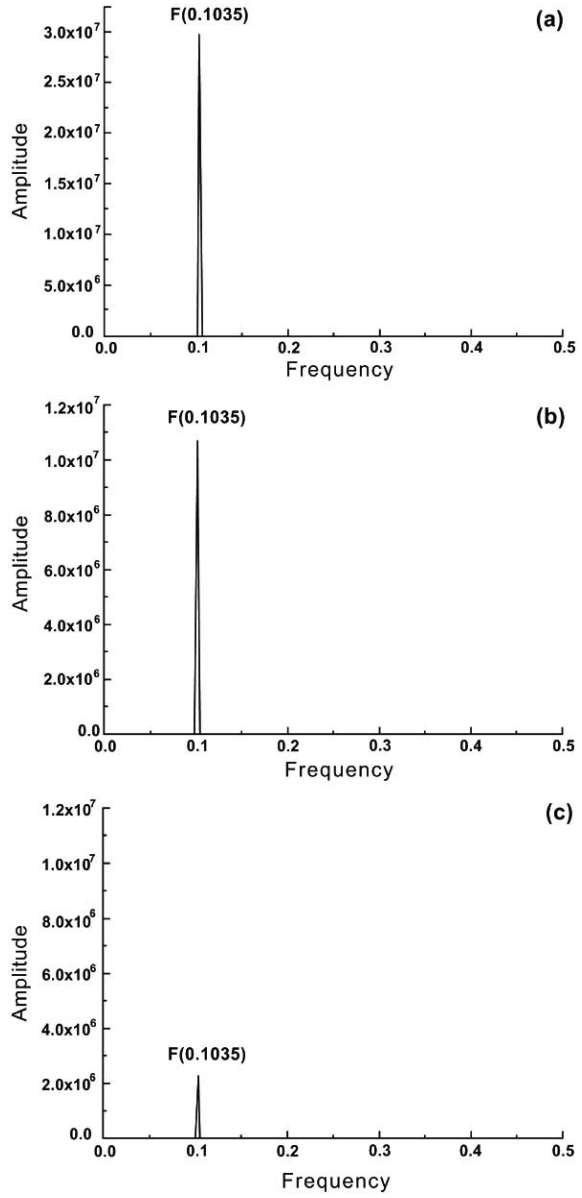

Fig. (2). Informational spectrum analyses of HIV-1 gp120. (a) The consensus informational spectrum of 50 gp120 from R5 HIV-1 isolates [74] and the cross-spectrum between CIS of gp120 and **(b)** the informational spectrum of the CCR5 and **(c)** CXCR4 receptor. For each spectrum the abscissa represents the frequencies from the Fourier transform of the sequence of the electron-ion interaction potential corresponding to the amino-acid sequence of the protein. The lowest frequency is 0.0 and the highest is 0.5. The ordinate represents amplitudes, in arbitrary units, corresponding to each frequency component in the informational spectrum.

In order to prove this theoretical prediction, binding experiments *in vitro* using a recombinant protective antigen and EMILIN-1 have been performed. Results of these experiments presented in Fig. (**5**) confirmed strong binding between these two proteins as predicted by the cross-spectral analysis.

The formal demonstration that protective antigen interacts also with EMILIN-1 *in vitro* deserves some consideration. One possibility is that EMILIN-1, which is particularly rich in vascular tissues, could represent an elective site for *Bacillus Anthracis* deposition. This possibility is in accord with the finding that the human inhalational anthrax causes vascular lesions which are responsible for the pronounced hemorrhages.

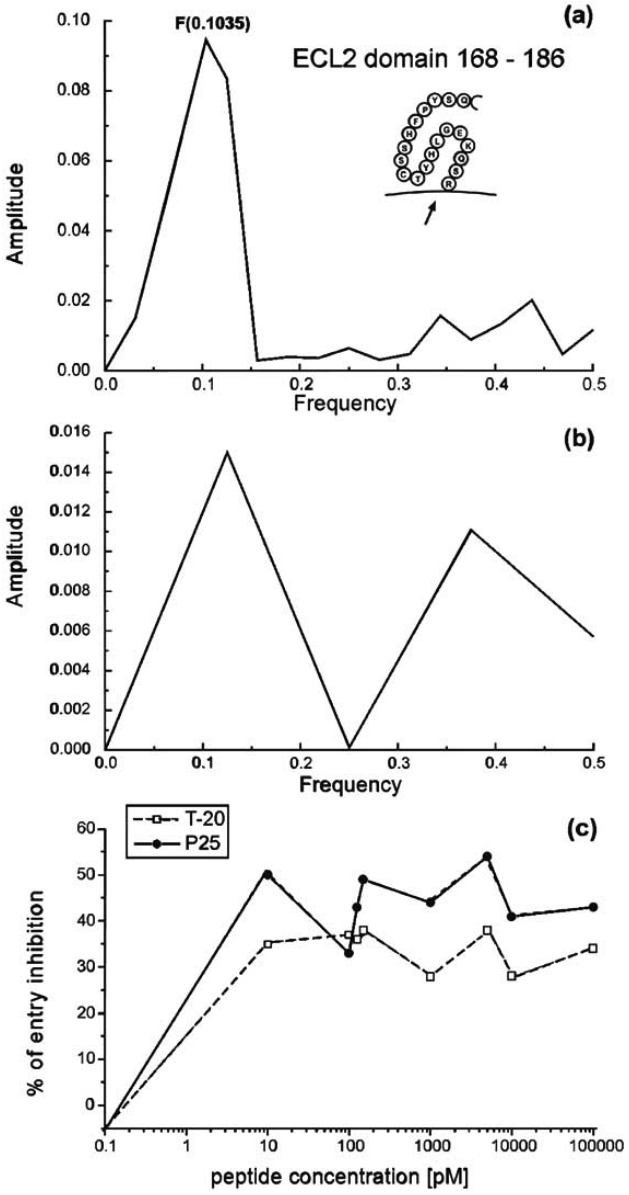

Fig. (3). HIV-1 peptide inhibitor discovery by ISM and the phage library screening. (a) The informational spectrum of the second extracellular loop (ECL2) of CCR5 receptor encompassing residues 168 – 186. **(b)** The informational spectrum of peptide P25 selected by screening of the phage library with ECL2-derived peptide 168-186. The abscissa and ordinate are as described in Fig. (**2**). **(c)** Entry inhibition of R5 HIV-1 isolate JR-FL with peptide P25 selected by screening of the phage library and the commercial HIV-1 entry inhibitor peptide T-20 (Fuzeon) [74].

Assessment of Biological Effects of Mutations as a Risk Factor for Cardiovascular Diseases [76]

Lipoprotein lipase (LPL) plays a central role in lipoprotein metabolism and energy homeostasis of all vertebrates.

Decrease of the LPL enzymatic activity caused by mutations leads to elevated triglycerides and reduced high-density lipoprotein levels, both risk factors for cardiovascular disease.

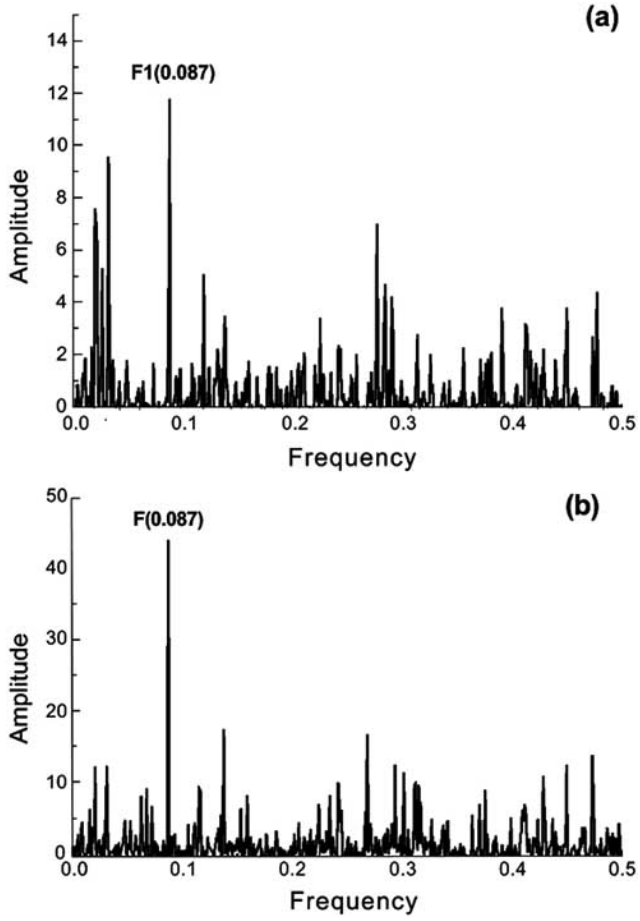

Fig. (4). Informational spectrum analyses of PA, its receptor and EMILIN-1. (a) The cross-spectrum between the protective antigen of Bacillus Anthraces and the anthrax receptor CMG2. **(b)** The cross-spectrum between the protective antigen and human EMILIN-1 protein. The abscissa and ordinate are as described in Fig. (**2**).

Multiple sequence alignment of 14 vertebrate sequences revealed the overall homology of 43.3%, which is close to the homology of the randomly selected sequences (37%). As can be seen from results presented in Fig. (**6**), the CIS of analyzed LPL molecules contains two dominant peaks corresponding to major common frequency components F1(0.033) and F2(0.168). The presented result suggests that primary structures of LPL molecules from vertebrates, despite of limited homology, encode common information which is evolutionary conserved and which relates to the biological function of LPL. According to the ISM concept, mutated LPL molecules with decreased biological activity should have lower amplitudes at characteristic frequencies F1 and F2 compared to the wild type. In Table **2** are given 93 naturally occurring mutations in the human LPL molecule, which either completely or partially abrogate its enzymatic function. Amplitude values corresponding to at least one characteristic frequency component, F1(0.033) and

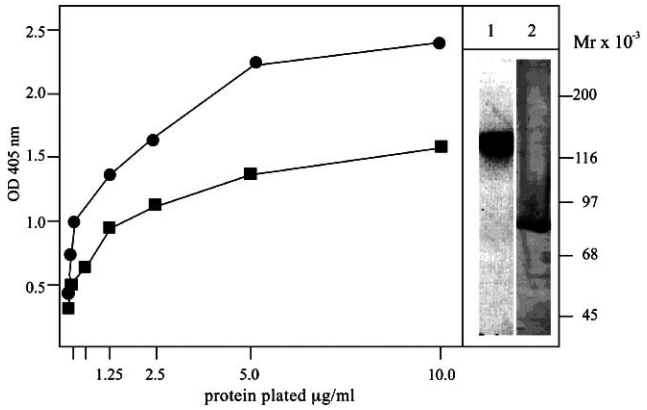

Fig. (5). Interaction between the protective antigen (PA) and EMILIN-1 [75]. Solid phase binding assays. Recombinant PA or EMILIN-1 were coated to the plastic surface of multiwell plates at serial dilutions, in concentrations as indicated and incubated with 10 μg/ml of EMILIN-1 (closed circles) or PA (closed squares), respectively. The data are the mean of triplicate assays. To the right the purification of PA and EMILIN-1 attained. The proteins were analyzed on 10% SDS-PAGE and stained with Coomassie blue. Line 1, EMILIN-1; Line 2, PA.

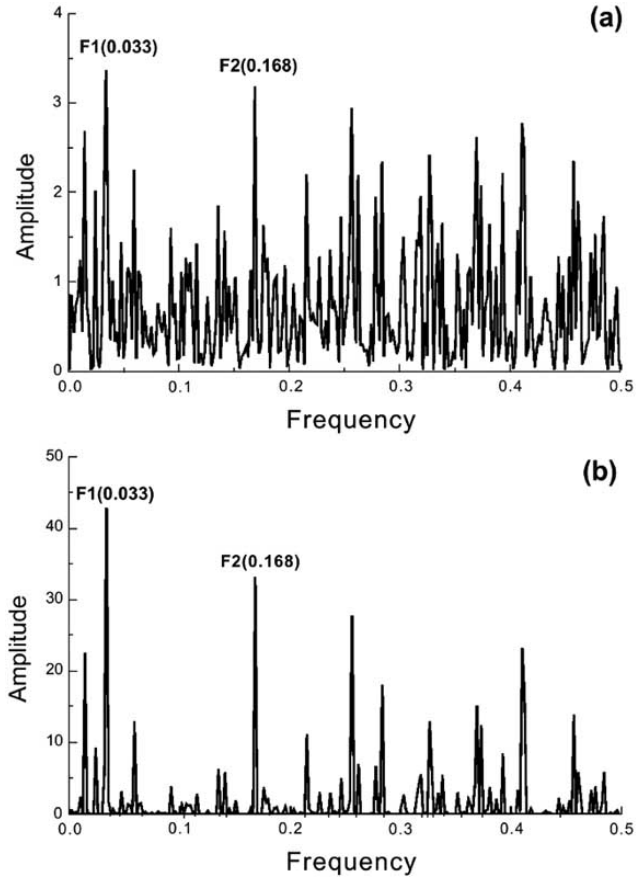

Fig. (6). Informational spectrum analyses of LPL. (a) The informational spectrum of the human LPL and **(b)** cross-spectra of the vertebrate LPL molecules. The abscissa and ordinate are as described in Fig. (**2**).

F2(0.168), are decreased in 75 of 93 analyzed mutant proteins (80.6%) (Table **2**). Further, 38 of 45 mutations (84.4%)

Table 2. Changes of Amplitudes at Frequencies F(0.033) and F(0.168) in Informational Spectrum of LPL Caused by Mutations

Mutation	F(0.033)	F(0.168)	Mutation	F(0.033)	F(0.168)	Mutation	F(0.033)	F(0.168)
D9N	+	+	R170L	+	+	L252R*	-	+
D21V	-	+	S172C	0	0	S259G	+	+
C27X	-	-	D174V	-	-	S259R*	-	-
N43S *	-	+	A176T*	+	-	A261T	-	-
H44Y	-	+	D180 E*	-	-	Y262H*	-	-
Y61X*	-	-	V181I	-	-	Y262X	-	-
W64X	-	-	H183D*	+	-	C264X	-	-
V69L	+	+	H183N*	-	+	S266P	-	+
A71T	-	-	H183Q	+	-	F270L*	-	-
Y73X	-	-	G188E*	-	-	C283Y	+	+
R75S	+	+	G188R	-	-	L286P	+	-
W86G	-	-	R192Q	+	-	Y288X*	-	-
W86R*	+	+	S193R	-	+	N291S	-	-
A98T	-	+	G195E*	-	-	S298R	+	+
T101A	+	+	D204E*	-	-	M301R*	+	-
G105R*	-	-	I205S*	+	+	M301T	+	-
Q106X*	-	-	P207L*	-	+	Y302X*	-	-
S132A*	+	-	C216S*	0	0	L303P*	+	-
H136R	-	-	I225T	-	-	L303F*	+	-
G139S*	-	+	C239X*	-	-	S323C	0	0
G142E*	+	-	C239W*	-	-	A334T*	+	+
G154S*	+	-	E242K	+	-	S338F	-	-
G154V*	+	-	R243C	-	+	T352I	-	-
D156G*	+	-	R243H*	-	+	L365V*	+	+
D156G*	+	-	R243L	-	+	W382X*	-	-
D156H*	-	-	S244T*	+	+	W382X	-	-
D156N*	+	-	I249T	-	-	E410V	0	0
P157R*	-	+	D250N*	+	-	E410K	-	-
A158T	-	-	S251C*	0	0	C418Y	-	-
E163D	-	+	L252P*	-	+	E421K	+	-
E163G	+	-	L252V*	-	+	C438S	0	0

* (mutations which completely abrogate the LPL enzyme activity)
- (decrease of the amplitude on characteristic frequency)
+ (increase of the amplitude on characteristic frequency)
0 (without change of the amplitude on characteristic frequency)

resulting in complete enzymatic LPL deficiency decrease in amplitude at one or both characteristic frequencies. These results point out the change of amplitudes at the IS frequencies F1(0.033) and F2(0.168), as a reliable bioinformatics criterion for assessment of the biological effects of LPL mutations which can be a useful tool in prevention of cardiovascular diseases.

ISM Analysis of Hemagglutinin from Highly Pathogenic Influenza Virus H5N1

The main obstacle in development of an effective vaccine for the influenza virus represents the high variability of its antigenic determinants. In order to solve this problem, we applied some experience from our HIV-1 research.

As we demonstrated above, despite high variability HIV-1 gp120 preserves its tropism to the CCR5 receptor and this property corresponds to specific information encoded by the primary structure of gp120 which is represented by the IS frequency F(0.1035) [74]. It was hypothesized that by analogy hemagglutinin (HA) of the influenza virus, which also is highly variable, encodes some conserved information which is responsible for its interaction with the receptor. In Fig. (**7**) the individual informational spectra and CIS of HA molecules from 3 reference H5N1 isolates are presented. Fig. (**8**) shows CIS of HA from highly virulent avian H5N1 viruses, randomly selected H5N2 viruses and H5N1 viruses isolated from infected patients. As can be seen, all CIS presented in Fig. (**7**) and Fig. (**8**) contain only one characteristic peak which corresponds to the IS frequency F(0.076). These results point out that primary structures of HA molecules from the H5 influenza subtype encode common information which is represented by the frequency component F(0.076).

In order to prove the specificity of this information cross-spectral analysis of HA molecules of other influenza subtypes has been performed. Results of this analysis are given in Fig. (**9**). As can be seen, the hemagglutinin molecule of each influenza subtype encodes specific information which is represented by different frequency components. This information for H5 HA corresponds to the IS frequency F(0.076).

In Fig. (**10**) are compared CIS of HA molecules from the highly pathogenic "Spanish flu" H1N1 isolates with other nonpathogenic H1N1 isolates and with highly pathogenic H5N1 viruses. Significant difference between consensus spectra of highly pathogenic "Spanish flu" isolates and other nonpathogenic H1N1 isolates is evident, although these viruses belong to the same subtype of the influenza virus. It is important to note that the consensus spectrum of "Spanish flu" isolates also contains the frequency component F(0.076) which is characteristic for the H5 subtype. It means that the primary structure of HA antigens from "Spanish flu" isolates and H5N1 viruses encode information which probably determines a common biological property of these molecules. This result indicates that, from the informational point of view, the "Spanish flu" HA antigen could represent a precursor of H5 HA.

The main contribution to information corresponding to the IS frequency F(0.076) gives the domain (denoted VIN1) located in the N-terminus of H5 HA (Fig. (**11**)). In the tertiary structure of the HA molecule, the peptide VIN1 is located in the site E (Table **3**), one of the major antigenic sites of this molecule. This antigenic site is located below the globular head encompassing the receptor binding site [82]. It is of note that most of the escape mutants have mutations in this globular part of the HA molecule. These mutations allow the virus change of the receptor tropism and avoid the immune response. On the other hand, escape mutants carrying

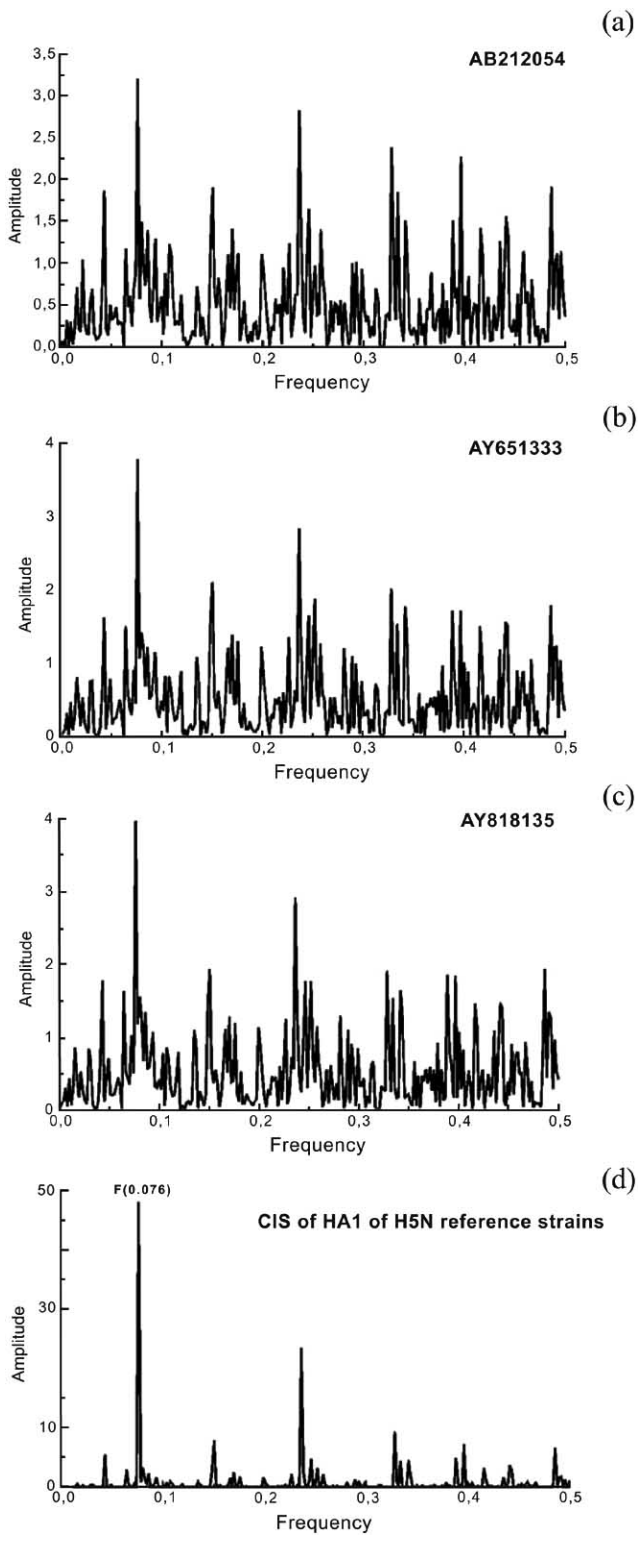

Fig. (7). Informational spectrum analyses of HA1 of the reference H5N1 avian influenza viruses (**AIV) isolates. (a)** The informational spectra of subunit 1 of HA (HA1) of the reference H5N1 AIV isolates: A/Hong Kong/213/2003(H5N1), A/Viet Nam/1194/2004(H5N1) and A/Viet Nam/1203/2004(H5N1) (GenBank Acc. No. AB212054, AY651333, AY818135, respectively) and **(b)** their consensus spectrum. The abscissa and ordinate are as described in Fig. (**2**).

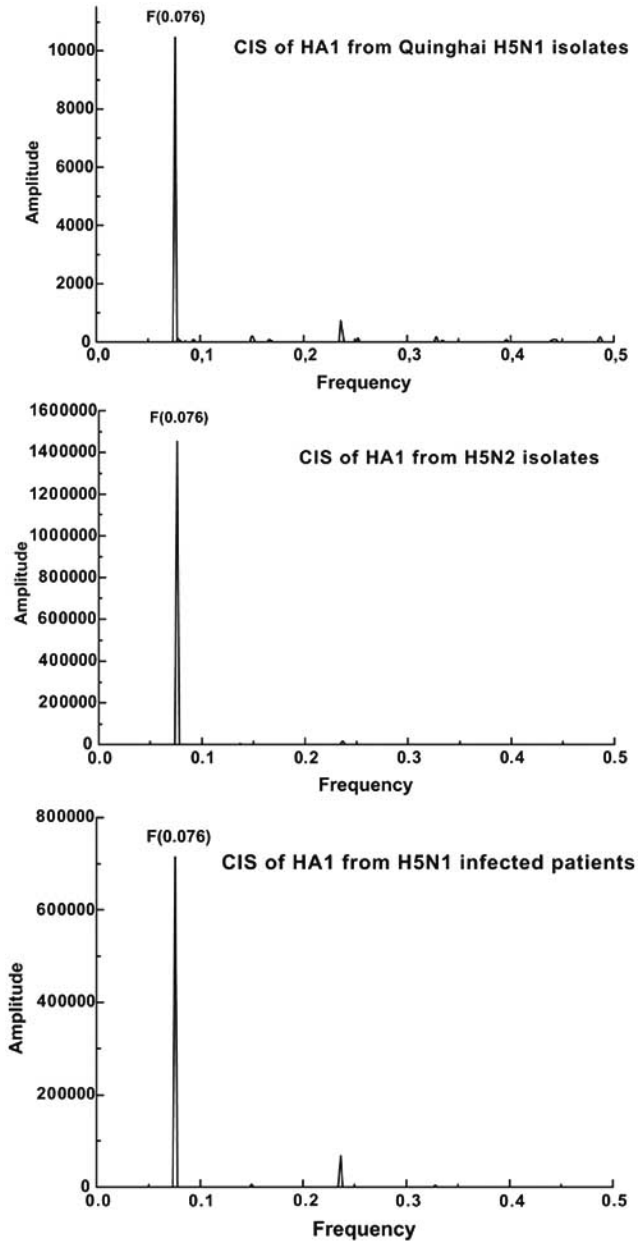

Fig. (8). CIS of HA from highly virulent avian H5N1 viruses, randomly selected H5N2 viruses and H5N1 viruses isolated from infected patients contain only one characteristic peak. (a) CIS of HA1 of the high virulent of H5N1Quinghai isolates (GenBank Acc. No. DQ095630, DQ095631, DQ095626, DQ095623, DQ095622, DQ095620, DQ095619, DQ095614). **(b)** CIS of HA1 of randomly selected H5N2 isolates (A/Turkey/Virginia/6962/83(H5N2),A/Chicken/Virginia/40018/84(H5N2), A/Chicken/New Jersey/12508/86(H5N2), A/Chicken/Florida/27716-2/86(H5N2), A/Chicken/New York/12004-3/87(H5N2), A/Chicken/Michoacan/28159-530/95 (H5N2), A/Chicken/ Mexico/28159-541/95, (H5N2), A/Chicken/Jalisco/28159-600/95 (H5N2), A/Chicken/ Guatemala/45511-2/00 (H5N2), A/Chicken/El Salvador/102711-1/01 (H5N2)). **(c)** CIS of HA1 of H5N1 strains isolated from patients (GenBank Acc. No. AJ715872, AY651335, AF102671, AY555153, AY627885, AY651333, AY651336, AY651334, AF046088, AY575869, AB212054). The abscissa and ordinate are as described in Fig. (**2**).

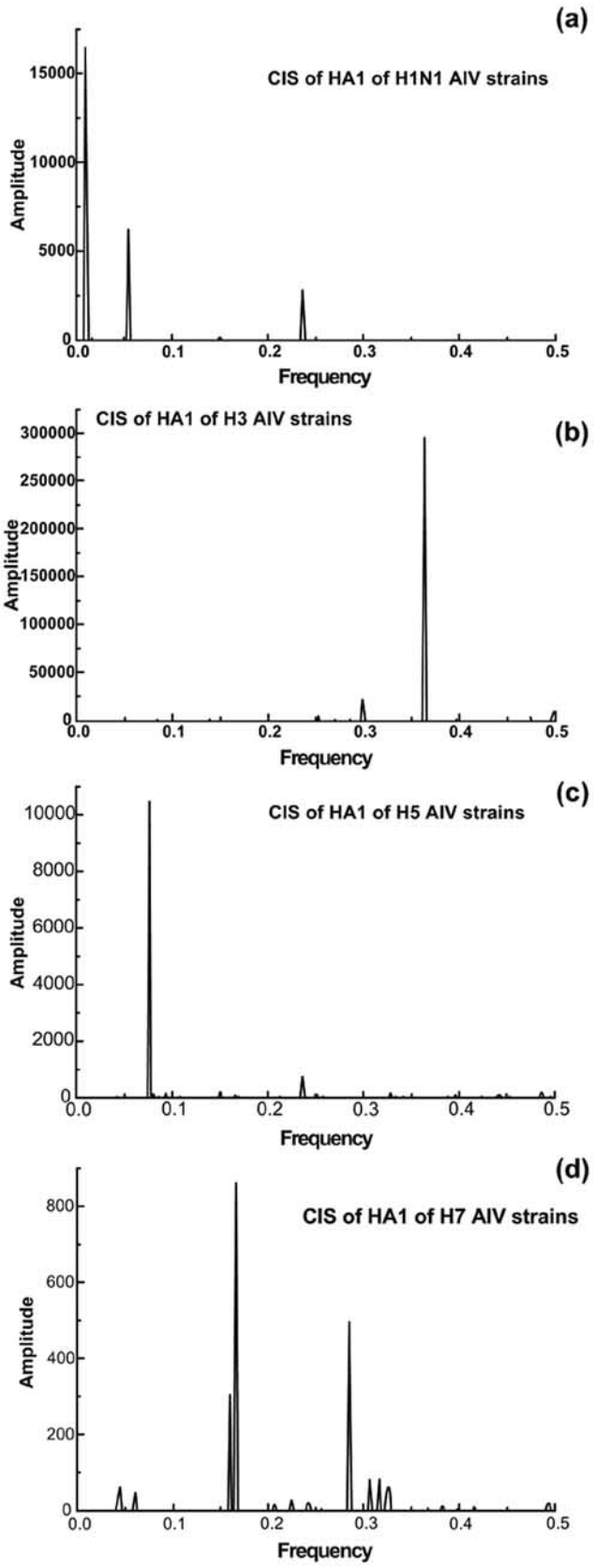

Fig. (9). Informational spectrum characteristics of HA1. The consensus informational spectra of HA1 of **(a)** H1, **(b)** H3, **(c)** H5 and **(d)** H7 AIV. The abscissa and ordinate are as described in Fig. (**2**).

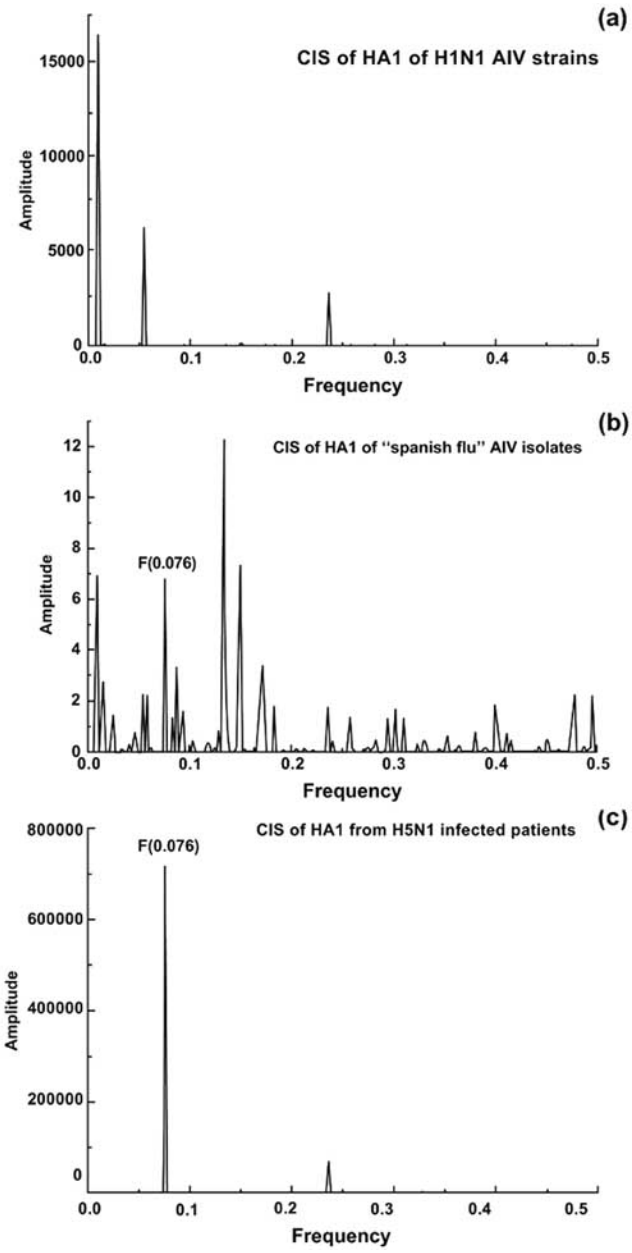

Fig. (10). Common informational characteristics of "Spanish flu" pandemic isolates and H5N1 viruses isolated from infected patients. The consensus informational spectra of HA1 of **(a)** H1N1 AIV isolates, **(b)** "Spanish flu" pandemic isolates (A/South Carolina/1/18 (H1N1), A/New_York/1/18(H1N1), A/Brevig_Mission/1/18(H1N1) and sequence reported in [83]) and **(c)** H5N1 viruses isolated from infected patients. The abscissa and ordinate are as described in Fig. (**2**).

mutations in the site E, are very rare. This structural arrangement indicates that variable antigenic sites A and B represent an immune decoy which protects sensitive site E containing the strongest T cell epitope. Similar structure is reported for gp120 in which the N-terminus of the second conserved domain, containing the strongest T cell epitope is protected by the hypervariable V3 loop, representing the principal antigenic determinant of HIV [80].

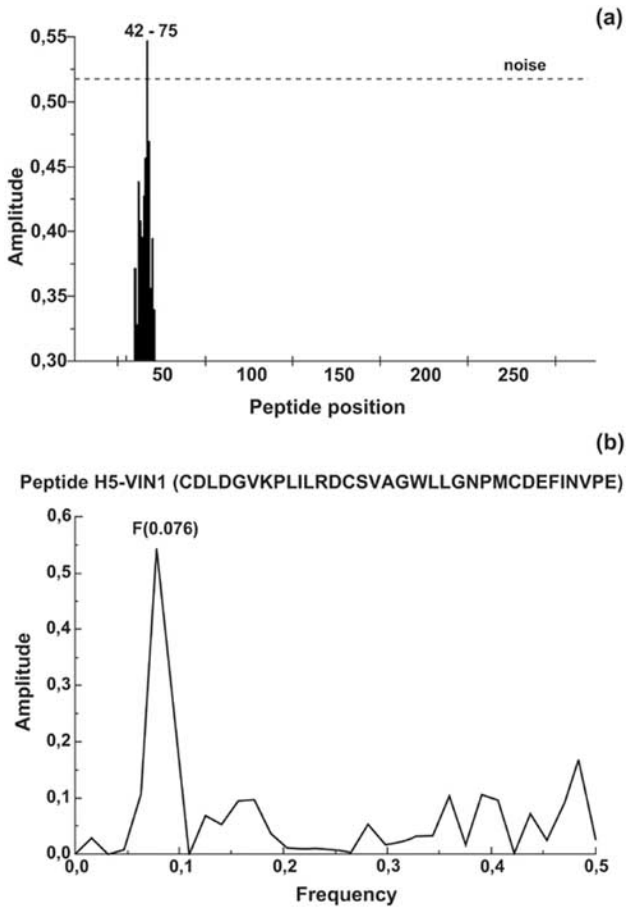

Fig. (11). Mapping of HA1 H5N1. (a) Results of mapping of HA1 of H5 reference influenza isolate A virus (A/Hong Kong/213/03(H5N1)) (GeneBank Ass. No. AB212054) with 34 a.a. peptide on the characteristic frequency F(0.076). The noise level corresponds to IS of the native HA1 molecule. **(b)** The informational spectrum and position of the peptide VIN1 in the primary structure of the influenza A virus (A/Hong Kong/213/03(H5N1)) HA1. The abscissa and ordinate are as described in Fig. (**2**).

Table 3. HA Monomer Major Antigenic Sites [82]

Antigenic site of epitope on the HA molecule	Amino acids residues
A	143, 144
B	188, 189
D	205, 218, 226
E	60, 63

In conclusion, (i) the primary structures of H5 HA molecules encode specific and highly conserved information which can be used for selection of antigens for universal anti-H5 vaccines which could be resistant to the escape mutants, and (ii) that the conserved peptide VIN1, derived from the N-terminus of H5N1 hemagglutinin comprising residues 42 to 75, represents a candidate immunogen for the diagnos-

tic ELISA test and anti-H5N1 vaccine which would be resistant to escape mutants.

DISCUSSION

A number of strategies have been proposed and used for rational discovery of new therapeutic drugs. All of these strategies focus on design of lead compounds having biological activity against a defined protein or nucleic acid target. For this reason, identification of a therapeutic target represents an important early step of the lengthy and costly process of drug development. A computer drug target identification strategy is feasible if a sufficiently diverse set of 3D structures of protein and nucleotide sequences is available. Unfortunately, these data currently are available only for a tiny fraction of human and microbial proteins which are accessible in protein and nucleotide databases. Partially because of this limitation, the total number of currently available molecular drug targets (human genome-derived as well as viral or microbial targets) is 483 [81]. Compared with 100.000 genes, a popular estimate for the total number of human genes, and over four million protein sequences in the sequence databanks, this number of therapeutic targets represents a very small fraction. This limitation does not concern ISM which can be applied to all protein and nucleotide sequences which are available in databases, because it uses as input information only the primary structure of biological macromolecules. As we demonstrated here, the ISM approach can be used for: (i) structure-function analysis of proteins and identification of functionally important domains representing candidate therapeutic targets, (ii) creation of the *in silico* interactomes and searching databases for possible interactors of selected proteins and nucleotide sequences, and (iii) assessment of biological effects of mutations. There are more than 70 groups worldwide (Table **4**) that have used EIIP/ISM approach in analysis of proteins and DNA sequences.

Finally, we would like to point out some limitations of ISM. This bioinformatics approach is not applicable for analysis of proteins whose biological function does not include long-range protein-protein interaction (e.g. structural proteins, transport proteins, etc.). The cross-spectral analysis of a large number of proteins usually gives a consensus spectrum which contains only one peak corresponding to the dominant common frequency component. For this reason, other common frequency components which are represented by lower amplitudes in the informational spectra of analyzed proteins are "screened" by this dominant frequency component. A typical example represents the ISM analysis of HIV-1 gp120. As we previously demonstrated [33], the CIS of gp120 from macrophage (M)-tropic and T cell (T)-tropic HIV-1 viruses contains only one peak at the frequency F(0.031). This frequency represents the common information which corresponds to the interaction between gp120 and the CD4 receptor. The primary structures of HIV-1 gp120 from M-tropic isolates encode common information corresponding to interaction with the CD4 receptor and the CCR5 coreceptor. The information are represented by the IS frequencies F(0.035) and F(0.1035), respectively. Because amplitudes at F(0.1035) are remarkably higher than amplitudes at F(0.035) in the IS of gp120 from the M-tropic HIV-1 strains, the CIS of these proteins contains only one peak correspond-

Table 4. Groups that Use EIIP/ISM Approach

Australia
Australia
• Department of Electrical and Computer Systems Engineering, Monash University, Caulfield • School of Electrical and Computer Engineering, RMIT University, Melbourne • The University of New South Wales, Sydney • Bioinformatics Applications Research Centre, James Cook University, Townsville
Asia
Japan
• Musashigaoka College, Yoshima, Higi-gun, Saitama • Nagoya City University School of Medicine, Mizuho-cho, Nagoya • BioFrontier Institute Inc., Nishi-Hashimoto 5-4-21, Sagamihara • Tokyo University of Technology, Tokyo • Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto • Human Genome Center, Tokyo University, Minatoku, Tokyo • Performance Evaluation Laboratory, University of Aizu, Aizu • Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto • Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo
New Zealand
• Department of Information Science, University of Otago, Dunedin,
India
• Chemical Engineering Division, National Chemical Laboratory, Pune • College of Engineering Chengannur, Kerala • Centre for Bioinformatics, University of Kerala, Kerala • Res & Training Unit Navigat Elect, Osmania Univ, ,Hyderabad
Singapore
• Biomedical Science Institute, Biopolis, Singapore • Institute for Infocomm Research, Singapore
United Emirates
• U.A.E. University, Al-Ain, United Emirates
China
• Department of Biology, Fundan University, Fundan • Joint Research Center for Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong • College of Chemistry, Sichuan University, Chengdu • State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan University, Changsha • Bioinformatics Division, TNLIST and Department of Automation, Tsinghua University, Beijin • College of Information Sciences and Technology, Donghua University, Shanghai • College of Sciences, Shanghai University, Shanghai
South Korea
• Department of Life Science, Pohang University of Science and Technology, Pohang
Taiwan
• Department of Electrical Engineering, National Chung Hsing University, Taichung • Department of Industrial Engineering and Management, Yuan Ze University, Chungli City
South America
Brazil
• Instituto de Matem´atica e Estat´ıstica, Universidade de Sao Paulo, Sao Paulo

(Table 4 contd…)

North America
USA
• Department of Electrical Engineering, California Institute of Technology, Pasadena • School of Molecular Biosciences , Washington State University, Pullman • Department of Electrical & Computer Engineering, University of Delaware, Delaware • R&D Division, DuPont Company, Wilmington • Rensselaer Polytechnic Institute, Troy • Delaware Biotechnology Institute, University of Delaware, Delaware • Department of Plant Pathology, Kansas State University, Manhattan • Division of Engineering, Brown University • Department of Physics/Engineering, Philadelphia Community College, Philadelphia • Thayer School of Engineering, Dartmouth College, Dartmouth • National Center for Biotechnology Information, National Library of Medicine, Bethesda • Dept. of Computer and Information Sci. and Eng., University of Florida, Gainesville • Department of Electrical Engineering , Arizona State University, Phoenix • Gordon Life Science Institute, San Diego • Dartmouth College Computer Science, Hanover • Department of Electrical Engineering, Arizona State University, Phoenix
Canada
• Department of Electrical and Computer Engineering, University of Victoria, • Dept Elect & Comp Engn, Concordia Univ, Montreal
Africa
South Africa
• Centre for Engineering Research, Technikon Natal, Durban • South African National Bioinformatics Institute (SANBI), Durban • Department of Electronic Engineering, University of Natal, Durban
Europe
Finland
• Tempere University of Technology, Tampere
Italy
• International Center for Genetic Engineering and Biotechnology, Trieste • Università degli Studi di Genova, Genova
Switzerland
• Molecular Design and Bioinformatics, Pharm. Div., F. Hofmann-La Roche Ltd., Basel
Greece
• Department of Cell Biology and Biophysics, University of Athens, Athens
England
• Dept. of Biochemistry and Molecular Biology, University College London, London, • Department of Computer Science, University College London, London • Edward Jenner Institute, Compton, Newbury, Berkshire • Department of Computing and Centre for BioMedical Informatics, University of Kent, Canterbury • Departments of Computer Science and Electronics, University of York, Heslington • Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, • Intelligent and Distributed Systems Laboratory, Liverpool Hope University, Liverpool • Medivir UK Ltd., Little Chesterford
Germany
• Dept. for Simulation of Biological Systems, Eberhard Karls University, Tübingen • AG Angewandte Informatik, Technische Fakultat, Universit at Bielefeld, Bielefild
Croatia
• Institute Rudjer Boskovic, Zagreb • Department of Electronic Systems and Informatics, Faculty of Electrical and Computer Engineering, • University of Zagreb, Zagreb
France
• Lab. de Biochimie et Génétique MoléculaireUniversité de La Réunion, Saint Denis • Institute Jacqe Monod, Paris • Laboratoire de Physique et Chimie Biomoleculaire, Universite Paris, Paris
Serbia
• Center for Multidisciplinary Research, Institute of Nuclear Sciences VINCA

ing to frequency component F(0.1035). Finally, it is of note that a common frequency component in the informational spectra of two proteins is not enough for their interaction, because their chemical binding is determined by their structural properties. On the other hand, only structural complementarity between interacting proteins is also not enough for their efficient interaction. Only both frequency and structural complementarities represent a necessary and sufficient condition for efficient protein-protein interaction.

As a corollary, ISM represents a simple and efficient bioinformatics tool for: (i) structure-function analysis of proteins and identification of functionally important domains representing a candidate therapeutic target, (ii) creation of the *in silico* interactomes and searching databases for possible interactors of selected proteins and nucleotide sequences, and (iii) assessment of biological effects of mutations.

FURTHER DEVELOPMENTS AND PERSPECTIVE

An important initial step of expensive and time-consuming process of drug development represents identification of therapeutic targets, entailing analyzing the causes of the diseases and finding ways to tackle them. The ISM-based bioinformatics platforms, allowing fast and inexpensive screening of large protein and DNA databases, in silico creation of different protein interactoms and assessment of biological effects of mutations, could significantly accelerate this process. Also future development of new ISM platforms that will include different signal processing techniques could also significantly speed-up development of new drugs.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science of the Republic of Serbia (Contract No. 143001).

ABBREVIATIONS

3D	=	Three-dimensional
AIV	=	Avian influenza viruses
AQVN	=	Average quasivalence number
CIS	=	Consensus informational spectrum
CMG2	=	Capillary morphogenesis protein 2
ECL2	=	Second extracellular loop
EIIP	=	Electron-ion interaction potential
EMILIN	=	Elastin microfibril interfase located protein
gp120	=	Glycoprotein 120
HA	=	Hemagglutinin
HIV	=	Human immunodeficiency virus
IS	=	Informational spectrum
ISM	=	Informational spectrum method
LPL	=	Lipoprotein lipase
M-tropic	=	Macrophage-tropic
Ry	=	Rydbergs units
S/N	=	Signal-to-noise ratio
T-tropic	=	T cell-tropic

REFERENCES

[1] Smoluchowski, M.V. (**1916**) *Phys. Z, 17*, 129-68.
[2] Northrup, S.H. and Erickson, H.P. (**1992**) *Proc. Natl. Acad. Sci. USA, 89*, 3338-42.
[3] Sharp, K., Fine, R. and Honig, B. (**1987**) *Science, 236*, 1460-3.
[4] Wiegel, F.W. and DeLisi, C. (**1982**) *Am. J. Physiol., 243*, R475-R479.
[5] McCloskey, M.A. and Poo, M.M. (**1986**) *J. Cell. Biol., 102*, 88-96.
[6] Peters, R. (**2005**) *Traffic, 6*, 421-7.
[7] Brune, D. and Kim, S. (**1994**) *Proc. Natl. Acad. Sci. USA, 91*, 2930-4.
[8] Camacho, C.J., Kimura, S.R., DeLisi, C. and Vajda, S. (**2000**) *Biophys. J., 78*, 1094-105.
[9] Fröhlich, H. (**1968**) *Int. J. Quantum Chem., 2*, 641-9.
[10] Fröhlich, H. (**1970**) *Nature, 228*, 1093.
[11] Fröhlich, H. (**1975**) *Proc. Natl. Acad. Sci. USA, 72*, 4211-5.
[12] Veljkovic, V. (**1980**) *A Theoretical Approach to Preselection of Carcinogens and Chemical Carcinogenesis*. Gordon & Breach, New York.
[13] Veljkovic, V. and Lalovic, D. (**1976**) *Cancer Biochem. Biophys., 1*, 295-8.
[14] Veljkovic, V. and Slavic, I. (**1972**) *Phys. Rev. Lett., 29*, 105–7.
[15] Veljkovic, V. (**1973**) *Phys. Lett. 45A*, 41–2.
[16] Veljkovic, V. and Lalovic, D. (**1973**) *Phys. Lett.,45A*, 59.
[17] Veljkovic, V. and Lalovic, D. (**1977**) *Experientia, 33*, 1228-9.
[18] Veljkovic, V.and Lalovic, D. (**1978**) *Experientia, 34*, 1342-3.
[19] Ajdacic, V. and Veljkovic, V. (**1978**) *Experientia, 34*, 633-5.
[20] Veljkovic, V. and Ajdacic, V. (**1978**) *Experientia, 34*, 639-1.
[21] Veljkovic, V., Mouscadet, J., Veljkovic, N., Glisic, S., and Debyser, Z. (**2007**) *Bioorg. Med. Chem. Lett. 17*, 1226-32.
[22] Tintori, C., Manetti, F., Veljkovic, N., Perovic, V., Hayes, S., Witvrouw, M., Debyser, Z., Vercammen, J., Veljkovic V. and Botta, M. (**2007**) *J. Chem. Infor. Model., 47*, 1536-44.
[23] Veljkovic, V., Cosic, I., Dimitrijevic, B. and Lalovic, D. (**1985**) *IEEE Trans. Biomed. Eng., 32*, 337-41.
[24] Veljkovic, V. and Cosic, I. (**1987**) *Cancer Biochem. Biophys., 9*, 139-48.
[25] Lazovic, J. (**1996**) *Comp. Appl. Bio. Sci., 16*, 553-8.
[26] Cosic, I. (**1997**) The Resonant Recognition Model of Macromolecular Bioreactivity: Theory & Application. Birkhauser Verlag, Berlin.
[27] Cosic, I., Nesic, D., Pavlovic, M. and Williams, R. (**1986**) *Biochem. Biophys. Res. Commun., 141*, 831-8.
[28] Veljkovic, V. and Metlas, R. (**1988**) *Cancer Biochem. Biophys., 10*, 91-106.
[29] Skerl, V. and Pavlovic, M. (**1988**) *FEBS Lett., 239*, 141-6.
[30] Cosic, I. and Nesic, D. (**1987**) *Eur. J. Biochem., 170*, 247-52.
[31] Cosic, I., Pavlovic, M. and Vojisavljevic, V. (**1989**) *Biochimie, 71*, 333-42.
[32] Cosic, I., Vojisavljevic, V., and Pavlovic, M. (**1989**) *Int. J. Radiat. Biol., 56*, 179-91.
[33] Veljkovic, V. and Lalovic, D. (**1990**) *Biosystems, 23*, 311-6.
[34] Cosic, I. and Hearn, M. (**1991**) *J. Mol. Recognit., 4*, 57-62.
[35] Cosic, I., Hodder, A., Aguilar, M. and Hearn, M. (**1991**) *Eur. J. Biochem., 198*, 113-9.
[36] Veljkovic, V., Metlas, R., Raspopovic, J. and Pongor, S. (**1992**) *Biochem. Biophys. Res. Commun., 189*, 705-10.
[37] Cosic, I. (**1994**) *IEEE Trans. Biomed. Eng., 12*, 1101-14.
[38] Cosic, I., Drummond, A.E., Underwood J.R. and Hearn M. (**1994**) *Mol. Cell. Biochem., 130*, 1-9.
[39] Cosic, I. and Birch, S. (**1994**) *IEEE EMBS, 16*, 265.
[40] Nair, T., Tambe, S. and Kulkarni, B. (**1994**) *FEBS Lett., 346*, 273-7.
[41] Veljkovic, V., Johnson, E. and Metlas, R. (**1995**) *Cancer J., 8*, 308.
[42] Cosic, I. (**1995**) *Biotechnology (NY), 13*, 236-8.
[43] Birch, S., West, R. and Cosic, I. (**1995**) *Australas. Phys. Eng. Sci. Med., 18*, 197-207.
[44] Cosic, I. (**1996**) *Med. Biol. Eng. Comput., 34*, 139.
[45] Krsmanovic, V., Biquard, J., Sikorska-Walker, M., Cosic, I., Desgranges, C., Trabaud, M., Whitfield, J., Durkin J., Achour, A. and Hearn, M. (**1998**) *J. Pept. Res., 52*, 410-20.
[46] Fang, Q. and Cosic, I. (**1998**) *Australas. Phys. Eng. Sci. Med., 21*, 179-85.
[47] de Trad, H.C., Fang, Q. and Cosic, I. (**2000**) *Biophys. Chem., 84*, 149-57.
[48] Parbhane, R.V., Unniraman, S., Tambe, S.S. and Kulkarni, B.D. (**2000**) *J. Biomol. Struct. Dyn., 17*, 665-72.
[49] Murakami, M. (**2000**) *J. Protein Chem., 19*, 609-13.
[50] de Trad, C.H., Fang, Q. and Cosic, I. (**2002**) *Protein Eng., 15*, 193-203.
[51] Arrigo, P., Ivaldi, G. and Cardo, P.P. (**2002**) *In Silico Biol., 2*, 143-50.
[53] Veljkovic, N., Branch, D., Metlas, R., Prljic, J., Vlahovicek, K., Pongor, S.and Veljkovic, V. (**2003**) *J. Pept. Res., 62*, 158-66.
[54] Shepherd, A.J., Gorse, D. and Thornton, J.M. (**2003**) *Proteins, 50*, 290-302.
[55] Chen, J., Li, H., Sun, K. and Kim, B. (**2003**) *IEEE Signal Process. Mag., 20*, 16-26.
[56] Pirogova, E., Simon, G.P. and Cosic, I. (**2003**) *IEEE Trans. Nanobiosci., 2*, 63-9.
[57] Ivanovic, V., Demajo, M., Todorovic-Rakovic, N., Nikolic-Vukosavljevic, D., Neskovic-Konstantinovic, G., Krtolica, K., Veljkovic, V., Prljic, J. and Dimitrijevic, B. (**2004**) *Med. Hypotheses, 62*, 727-32.
[58] Vaidyanathan, P.P. and Yoon, B. (**2004**) *J. Franklin Inst., 341*, 111-35.
[59] Sauren, M., Pirogova, and Cosic, I. (**2004**) *Australas. Phys. Eng. Sci. Med., 27*, 174-9.
[60] Krishnan, A., Li, K.B. and Issac, P. (**2004**) *In Silico Biol., 4*, 133-48.
[61] Vaidyanathan, P.P. (**2005**) *IEEE Circuits Syst. Mag., 4*, 6-29.
[62] Wen, Z., Wang, K., Li, M., Nie, F. and Yang, Y. (**2005**) *Comput. Biol. Chem., 29*, 220-8.
[63] Guo, Y., Li, M., Wang, Wen, K.Z., Lu, M., Liu, L.and Jiang, L. (**2005**) *Acta Biochim. Biophys. Sin. (Shanghai), 37*, 759-66.
[64] Rakovic, D., Dugic, M. and Plavsic, M. (**2005**) *Mater. Sci. Forum., 494*, 513-8.
[65] Riaz, T., Li, K.B., Tang, F. and Krishnan, A. (**2005**) *In Silico Biol., 5*, 415-8.
[66] Huang, J., Tan, P.H., Li, K.B., Matsumoto, K., Tsujimoto, M. and Bay, B.H. (**2005**) *Int. J. Oncol., 26*, 607-13.
[67] Veljkovic, N. and Perovic, V. (**2006**) *Protein Pept. Lett., 13*, 807-14.
[68] Pham, T.D. (**2006**) *Lect. Not. Artif. Intell., 3930*, 595- 604.

[69] Pham, T.D. (**2006**) *Lect. Not. Artif. Intell., 4065*, 24-37.
[70] Guo, Y.Z., Li, M., Lu, M., Wen, Z., Wang, K. Li, G. and Wu, J. (**2006**) *Amino Acids*, *30*, 397-402.
[71] Pirogova, E., Vojisavljevic, V., Fang, Q. and Cosic, I. (**2006**) *Mol. Simul.*, *32*, 1195-203.
[72] Pham, T.D. (**2006**) *IEEE Trans. Nanobiosci.*, *5*, 83-8.
[73] Pham, T.D. (**2007**) *Pattern Recog.*, *40*, 516-29.
[74] Veljkovic, V., Veljkovic, N., Este, J.A., Huther, A. and Dietrich, U. (**2007**) *Curr. Med. Chem.*, *14*, 441-53.
[75] Doliana, R., Veljkovic, V., Prljic, J., Veljkovic, N., Mongiat, M., De Lorenzo, E. and Clolombatti, A. (**2008**) *Matrix Biol., 27*, 96-106.
[76] Glisic, S., Arrigo, P., Alavantic, D., Perovic, V., Prljic, J. and Veljkovic, N. (**2008**) *Proteins*, *70*, 855-62.
[77] Djordjevic, A., Veljkovic, M., Antoni, S., Sakarellos-Daitsiotis, M., Krikorian, D., Zevgiti, S., Dietrich, U., Veljkovic, N. and Branch, D. (**2007**) *Curr. HIV Res.*, *5*, 443-8.
[78] Scobie, H.M., Rainey, G.J., Bradley, K.A. and Young, J.A. (**2003**) *Proc. Natl. Acad. Sci. USA*, *100*, 5170–4.
[79] The Uni Prot Consortium (**2007**) *Nucleic Acids Res.*, *35* (Database issue), D193-D197
[80] Veljkovic, V. and Metlas, R. (**1992**) *Immunol. Lett.*, *31*, 11-4.
[81] Overington, J.P., Al-Lazikani, B. and Hopkins, A.L. (**2006**) *Nat. Rev. Drug Discov.*, *5*, 993-6.
[82] Anders, E.M., Kapaklis-Deliyannis, G.P. and White, D.O. (**1989**) *J. Virol.*, *63*, 2758-67.
[83] Taubenberger, J.K., Reid, A.H., Krafft, A.E., Bijwaard, K.E. and Fanning, T.G. (**1997**) *Science*, *275*, 1793-6.

Received: February 04, 2008 **Revised: June 02, 2008** **Accepted: June 09 2008**