Application Of The EIIP/ISM Bioinformatics Concept in Development of New Drugs

V. Veljkovic[*,1], N. Veljkovic[1], J. A. Esté[2], A. Hüther[3] and U. Dietrich[3]

[1]*Centre for Multidisciplinary Research, Institute of Nuclear Sciences VINCA, Belgrade 11001, Serbia*

[2]*Retrovirology Laboratory IrsiCaixa, Hospital Universitari Germans Trias i Pujol,Trias i Pujol, Badalona ES-08916, Spain*

[3]*Georg-Speyer-Haus, Institute of Biomedical Research, Frankfurt-am-Main D-60596, Germany*

Abstract: The development of a new therapeutic drug is a complex, lengthy and expensive process. On average, only one out of 10,000 – 30,000 originally synthesized compounds will clear all the hurdles on the way to becoming a commercially available drug. The process of early and full preclinical discovery and clinical development for a new drug can take twelve to fifteen years to complete, and cost approximately 800 million dollars. 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 reviewed the application of the EIIP/ISM bioinformatics concept for the development of new drugs. This approach, connecting the electron-ion interaction potential of organic molecules and their biological properties, can significantly reduce development time through (i) identification of promising lead compounds that have some activity against a disease by fast virtual screening of the large molecular libraries, (ii) refinement of selected lead compounds in order to increase their biological activity, and (iii) identification of domains of proteins and nucleotide sequences representing potential targets for therapy. Special attention is paid in this review to the application of the EIIP/ISM bioinformatics platform along with other experimental techniques (screening of a phage displayed peptide libraries, testing selected peptides and small molecules for antiviral activity *in vitro*) in development of HIV entry inhibitors, representing a new generation of the AIDS drugs.

Keywords: Drug development, virtual screening, EIIP/ISM concept, HIV-1 entry inhibitors

INTRODUCTION

The development of drugs is a complex, expensive and time-consuming process. This process is also risky because estimated final success rates at which new drugs pass through all development stages to the market vary for different therapeutic classes from 12% to 33% [1]. DiMassi and co-workers analyzed the research and development cost for 68 randomly selected drugs [2]. According to this analysis, the total R&D cost per new approved drug is 900 million dollars in 2000 (preclinical costs 335 million dollars, clinical costs 467 million dollars and approval costs 100 million dollars). The R&D costs for new drugs as a whole also have increased substantially over time. It has been estimated that the growth rate in inflation-adjusted capitalized costs is 7.4% [2]. It means that, if growth rates were maintained and R&D of a new drug is initiated in 2001 with approvals obtained 12 years later, the capitalized cost would exceed 2 billion dollars.

As Jeffrey Augen presented in his recent review, during the past 25 years, the drug discovery process and a variety of information technologies have co-evolved to the point where they have become inseparable components of a pipeline that begins with basic research and ends with disease specific pharmaceuticals. The impact of this trend has been tremendous with regard to acceleration of the drug discovery process [3]. According to results of some studies, the addition of computer-aided design techniques, including docking studies, cheminformatics and bioinformatics, to the R&D could lead to a reduction in the cost of new drug development by up to 50% [4].

ROLE OF LONG-RANGE INTERMOLECULAR INTERACTIONS IN BIOLOGICAL SYSTEMS

The drug discovery process consists of seven basic steps: (*i*) disease selection, (*ii*) target identification, (*iii*) lead compound identification, (*iv*) lead optimization, (*v*) preclinical trial testing, (*vi*) clinical trial testing, and (*vii*) farmacogenomic optimization. The results of the second and third steps of this process are especially important because they directly influence success and efficacy of all further steps which are costly and time-consuming. Target identification involves acquiring molecular level understanding of a specific disease state and includes analysis of gene sequences, protein structures and metabolic pathways. The third step in the drug discovery pipeline involves the identification of lead compounds as a complement to the target discovery process. Understanding how biomolecules recognize each other is very important to protein engineering and ligand (drug) design involved in these two steps of drug development. According to the current concept, success of this recognition mainly depends on surface complementarity between interacting biomolecules. Historically, this concept is based on two

*Address correspondence to this author at the Laboratory for Multidisciplinary Research, Institute of Nuclear Sciences VINCA, Mihaila Petrovica 12-14, 10001 Belgrade, Serbia; Tel: +381 11 2453 686; Fax: +381 11 3440 100; E-mail: vv@vin.bg.ac.yu

0929-8673/07 $50.00+.00

fundamental hypotheses: "key-and-lock" hypothesis and "induced fit" hypothesis. The first hypothesis was originally proposed by Emil Fischer in 1894 and was later extended by J. B. S. Haldane in 1930 who suggested that a degree of misfit between the enzyme and its substrate is needed to drive the chemical reaction forward. Paul Ehrlich, who had worked with Emil Fischer, regarded the specificity of the toxin-antitoxin and the antibody-antigen interaction as further examples of Fischer's 'key and lock' hypothesis. This concept was further extended by the "induced fit" hypothesis proposed by Daniel Koshland who suggested that the enzyme molecule changes shape as the substrate molecules approaches. The change in shape is 'induced' by the approaching substrate molecule. In other words, the protein structure is not viewed as rigid, and the substrate binding site is not viewed as being exactly complementary to the substrate. This more sophisticated model relies on the fact that molecules are flexible because single covalent bonds are free to rotate and it views the binding of the substrate. These two hypotheses served as a basis for development of different docking tools for the rational structure-based ligand design.

As can bee seen, methods for docking ligands in an active site have a long history, and the topic was well and frequently reviewed. The first docking tool DOCK [5] was designed based on the key-and-lock mechanism of protein–ligand recognition, considering both the ligand and the protein as rigid bodies. The next generation methods model the protein as rigid while allowing ligand flexibility [6-11]. The rationale behind this treatment is that small molecule ligands have fewer degrees of freedom than the protein, so it is computationally less expensive to handle them flexibly. Docking and screening tools further reached various levels of sophistication. Soft docking [12] handles protein flexibility implicitly by allowing a certain degree of interpenetration between the protein and the docked ligand, making the reasonable assumption that the exactly correct conformers of the protein and ligand are not sampled. The next level of sophistication is reached by using rotamer libraries [13,14] to sample the low energy conformations available to each side chain while optimizing the shape complementarity between the protein and the docked ligand [15-17].

The main goals for structure-based drug design focus on *de novo* custom designed lead compounds that have perfect charge and shape complementarity with the binding site on the target protein sequence. Unfortunately, this concept which is exclusively based on the short-range interactions (SRI) between drug and target is not accurate enough for predicting the potency of various lead compounds. We will describe here a new complementary concept which is based on the long-range interactions (LRI). Each of these two concepts is necessary but not sufficient for complete description of the protein-protein and target-drug interactions.

The described stereochemical complementarity together with the collision theory assuming that the first contact between interacting molecules is achieved accidentally by the thermal motions that cause molecular wander, represent a fundamental basis for our current understanding of intermolecular interactions in biological systems [for review, see Chap. 3 in Ref. 18]. According to this concept, before chemical bond formation takes place, reacting molecular regions must be positioned closed 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 (van der Waals, hydrogen bonding, ionic interactions, etc.). Many experimental observations are difficult to explain using such conceptualization because real biochemical processes proceed at rates which are a few orders of magnitude faster than the predicted ones [for review, see Ref. 19]. One of the most acceptable solutions for this problem was suggested by Froehlich [20-22]. Starting from the very general theoretical considerations he proposed that biological macromolecules (i) should have metastable excited states with a very high dipole momentum, and (ii) should be capable of strongly excited giant dipole vibrations with a frequency near 10^{11} Hz. The excitation of this polar state leads to the breaking of an electric symmetry, and through the arising internal electric field could have far-reaching biological consequences. Conclusive experimental evidence of the existence of periodic electric vibrations of the proteins in the range $10^{11} – 10^{12}$ Hz has been obtained by using the Raman effect and coherent millimeter waves [reviewed in Ref. 23]. The interactions between biological molecules in a polar medium (water and lipids), based on such vibrations lead to the appearance of frequency-selective long-range forces which are efficient at a distance longer than one linear dimension of the interacting macromolecules ($10^2 – 10^3$ Å) [20,23]. The concordance in frequencies between biological molecules allowing them further selective long-distance interactions will be referred to as *frequency complementarity*. The direct consequence of such complementarity between interacting molecules is an acceleration of biochemical processes by an increase of the number of productive collisions in comparison with the accidental encounter of molecules.

It has been proposed that the number of valence electrons and the electron-ion interaction potential (EIIP) representing the main energy term of valence electrons are essential physical parameters determining the long-range properties of biological molecules [24]. We showed [24,25] that EIIP for organic molecules can be determined by the following simple equation derived from the "general model pseudopotential" [26-28]:

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

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

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

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). The strong connection between EIIP and AQVN of organic molecules and their biological activity was demonstrated (mutagenicity, carcinogenicity, toxicity, antibiotic and cytostatic activity, etc.) [24, 29-33].

Informational Spectrum Method (ISM)

The EIIP parameter was also used as a basis for the informational spectrum method (ISM) for structure/function analysis of protein and nucleotide sequences [34,36]. The

ISM was successfully applied in analysis of different protein sequences, as well as in *de novo* design of biologically active peptides [34-77]. 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, determining electronic properties of amino acids and nucleotides, which are responsible for their intermolecular interactions. 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:

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

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 comprise 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 [34], which is defined as follows:

$$S(n) = X(n)X^{*}(n) = \tfrac{1}{2}X(n)\tfrac{1}{2},\ n = 1, 2, \ldots, N/2 \quad (4)$$

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 the 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 the informational spectrum (IS), representing the 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 frequencies in CIS are common frequency components for the analyzed sequences. A measure of similarity for each peak is the 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 the CIS for a group of proteins, which have different primary structures, and finds strictly defined peak frequencies, it means that primary structures of the analyzed proteins encode the same information. It was demonstrated that: 1) such a peak exists only for the group of proteins with the same biological function; 2) no significant peak exists for biologically unrelated proteins; 3) peak frequencies are different for different biological functions. Furthermore, it was shown that the proteins and their targets (ligand/receptor, antibody/antigen, etc.) have the same characteristic frequency in common [34]. Thus, it can be postulated that IS frequencies characterize not only the general function but also recognition and interaction between a particular protein and its target. Once the characteristic frequency for a particular protein function/interaction is identified, it is possible then to utilize the ISM approach to predict the amino acids in the sequence, which essentially contribute to this frequency and are likely to be crucial for the observed function.

EIIP/AQVN-Based Virtual Screening of CCR5 Antagonists

New technologies such as combinatorial chemistry and high-throughput screening (HTS) which allow synthesis of millions or possibly billions of compounds require chemists to confront an unimaginably large and diverse "chemical landscape". Attempts must therefore be made to introduce variety of computational techniques that allow chemists to reduce by virtual screening (VS) huge molecular libraries to a more manageable size. It has been estimated that *in silico* methods may significantly benefit drug development by saving an average of 130 million dollars and 0.8 years per drug [78].

There are two basic approaches to VS: (*i*) "VS by docking" which uses the 3D structure of the target protein binding site to prioritize compounds which possibly bind to the protein, and (*ii*) "similarity-based VS" which uses one or more compounds that are known to bind to the selected protein as a structural query. Both of these VS approaches are based on the structural complementarity and SRI between interacting molecules. Here we will apply a new approach of VS, which is based on AQVN and EIIP values of organic molecules determining their LRI properties, for selection of candidate HIV-1 entry inhibitors.

Despite advances in the development of potent anti-retroviral regimens that block HIV transcription and assembly, problems of drug resistance, latent viral reservoirs and drug induced toxic effects that compromise effective viral control point to the need for new classes of anti-HIV drugs with different modes of action. Viruses that are broadly resistant to currently available anti-retroviral medications also continue to represent a growing challenge for HIV therapy. Approximately 76% of HIV-patients with a measurable viral load are infected with a strain of the virus that is resistant to one or more classes of anti-retroviral agents, thus reducing treatment options. Hence, the identification of new antiretroviral drugs that have unique mechanisms of action and produce no or minimal adverse effects remains an important therapeutic objective. A new generation of antiviral drugs intended to counter HIV-1 entry

Table 1. The Training Set of CCR5 Antagonists Selected According to Recommendations Proposed in Ref. [78]

No.	Compound	Formula	IC_{50} [nM]	EC_{50} [nM]	AQVN	EIIP [Ry]
1	Aramaki2004 1	$C_{34}H_{41}ClN_2O_2$	1.4		2.500	0.0946
2	Aramaki2004 11b	$C_{32}H_{36}N_2O_4S$	200		2.720	0.0554
3	Aramaki2004 2	$C_{32}H_{35}NO_2$	950		2.571	0.0874
4	E913	$C_{28}H_41N_3O_4$	2	30	2.526	0.0926
5	E916	$C_{29}H_{41}N_5O_2$	7	70	2.520	0.0932
6	E917	$C_{29}H_{39}N_3O_3$	9	60	2.540	0.0912
7	Finke2001 1	$C_{29}H_{32}Cl_2N_2O_3S_2$	35		2.714	0.0985
8	Finke2001 3	$C_{25}H_{27}ClN_2O_4S_2$	10		2.852	0.0119
9	Finke2001 5	$C_{28}H_{32}Cl_2N_2O_2S$	30		2.597	0.0834
10	Finke2001 8bb	$C_{31}H_{39}ClN_2O_2S$	5		2.526	0.0926
11	GW873140	$C_{33}H_{44}ClN_3O_6$		0.5	2.621	0.0792
12	Hale2002 1	$C_{34}H_{37}ClN_4O_5$	0.8	1	2.765	0.0418
13	Hale2002 15	$C_{35}H_{50}N_2O_3$	2	40	2.422	0.0962
14	Hale2002 16	$C_{34}H_{48}N_6O$	0.2	4	2.472	0.0959
15	Imamura2004 4a	$C_{28}H_{33}N_3O$	18	64	2.554	0.0985
16	Imamura2004 4e	$C_{28}H_{32}ClN_3O$	5.9	9.4	2.554	0.0985
17	Imamura2004 4e	$C_{28}H_{31}ClFN_3O$	7.8	13	2.554	0.0985
18	Lynch 2003 1	$C_{32}H_{44}N_2O_2$	0.1	1.2	2.425	0.0963
19	Lynch 2003 32	$C_{32}H_{44}N_2O_2$	0.2	1.6	2.425	0.0963
20	SCH-C	$C_{28}H_{37}BrN_4O_3$		16	2.575	0.0868
21	SCH-D	$C_{28}H_{37}F_3N_4O_2$		0.5	2.486	0.0950
22	Seto2004 6o	$C_{37}H_{38}N_3O_2$	5.3	2.3	2.662	0.0703
23	Shah2005 1	$C_{32}H_{41}N_3O_4S$	2		2.642	0.0749
24	Shen2004 1	$C_{36}H_{48}N_4O_2$	1.2	2.4	2.489	0.0952
25	Shen2004 4	$C_{36}H_{47}FN_4O_2$	1.6	0.4	2.489	0.0952
26	Shen2004 2	$C_{30}H_{39}FN_2O_2$	4	100	2.460	0.0985
27	Tagat2004 21	$C_{33}H_{40}F_3N_5O$	1.4	0.35	2.512	0.0938
28	Tagat2004 3	$C_{27}H_{36}F_3N_5O$	2.8	0.39	2.472	0.0985
29	Tak-779	$C_{33}H_{39}N_2O_2$		2.0	2.540	0.0913
30	Tak-5h	$C_{30}H_{39}ClFN_3O_2$	0.29	0.23	2.474	0.0959
31	Tak-7a-diCl	$C_{29}H_{36}Cl_2FN_3O_2$	3.3	2.3	2.493	0.0950
32	Tak-7a	$C_{29}H_{39}N_3O_2$	16	72	2.493	0.0950
33	UK107,543	$C_{27}H_{30}N_4$	650	inactive	2.590	0.0845
34	UK179,645	$C_{34}H_{38}Cl_2N_4O$	1100	ND	2.557	0.0893
35	UK370,907	$C_{28}H_{31}N_3$	2	inactive	2.548	0.0903
36	UK372,673	$C_{27}H_{34}N_4O$	40	75	2.548	0.0903
37	UK374,503	$C_{27}H_{34}N_4O$	40	73	2.548	0.0903
38	UK 374,692	$C_{24}H_{34}N_4O$	49	inactive	2.548	0.0903

(Table 1). cotd.....

No.	Compound	Formula	IC_{50} [nM]	EC_{50} [nM]	AQVN	EIIP [Ry]
39	UK378,995	$C_{29}H_{36}N_4O$	2	11	2.543	0.0909
40	UK382055	$C_{29}H_{36}N_4O$	6	3	2.543	0.0909
41	UK395,859	$C_{30}H_{37}N_5O_2$	ND	1.8	2.622	0.0790
42	UK 396,794	$C_{30}H_{38}N_4O_2$	ND	0.6	2.568	0.0879
43	UK395,859	$C_{30}H_{37}N_5O_2$	ND	1.8	2.622	0.0790
44	UK 408,030	$C_{27}H_{39}N_5O$	7	9	2.472	0.0959
45	UK 418,526	$C_{28}H_{41}N_5O$	ND	2	2.453	0.0963
46	UK 427,857	$C_{29}H_{41}F_2N_5O$	ND	0.5	2.436	0.0964
47	UK 427,857-41	$C_{26}H_{36}F_3N_5O$	ND	14	2.451	0.0964
48	Willoughby2003 1	$C_{41}H_{49}NO_2$	0.8	50	2.473	0.0959
49	Willoughby2003 4A	$C_{30}H_{42}FN_5O_2$	1.8	11	2.500	0.0946
50	Willoughby2003 4c	$C_{31}H_{42}FN_5O_2$	1.1	11	2.518	0.0933
51	Willoughby2003 4A	$C_{30}H_{41}F_2N_5O_2$	0.4	33	2.500	0.0946

into susceptible cells is now being developed. The HIV-1 coreceptors are particularly attractive from the perspective of identifying new antiviral compounds, since they are seven-transmembrane-domain G protein-coupled receptors, a family of proteins that is a well validated target for drug development. In particular, the CCR5 coreceptor is essential for viral transmission and replication during the early, clinically latent phase of the disease, but also during the late-stage disease [79]. For this reason, the identification of inhibitors of CCR5-mediated HIV-1 fusion and entry has been the focus of antiviral drug development in recent years.

In order to establish an AQVN/EIIP criterion for selection of HIV-1 entry inhibitors by VS of molecular libraries we selected the training set (Table **1**) using the following guidelines proposed by Li and co-workers [80]: (i) a minimum of 16 diverse compounds should be selected to avoid any chance correlation, (ii) the activity data should have a range of 4 – 5 orders of magnitude, (iii) the compounds should be selected to provide clear and concise information and to avoid redundancy and bias in terms of both structural features and activity range, (iv) the most active compounds should be included so that they provide information on most critical features required as a pharmacophore, and (v) inclusion of any compound known to be inactive because of steric hindrance must be avoided. Table **1** shows the training set of CCR5 inhibitors selected according to these recommendations. It can be seen that 82.4% of all compounds and 90% of the most active CCR5 inhibitors (IC50 < 10nM) from this training set have EIIP and AQVN values within intervals (0.079 – 0.099 Ry) and (2.42 – 2.63), respectively. We will further use these EIIP and AQVN intervals as a criterion in VS of molecular libraries for candidate CCR5 inhibitors. It has previously been shown that the region of homogenous distribution of compounds in nature encompasses EIIP values between 0 and 0.13 Ry, and AQVN values in interval 2.0 - 3.8 [24].

In order to validate the proposed criterion we applied it in VS of the NIH molecular libraries of CCR5 inhibitors and HIV-1 integrase inhibitors. As can be seen from results presented in Table **2**, 65.61% of compounds from the NIH library of CCR5 inhibitors, which was used as the positive control, fit the proposed EIIP/AQVN criterion. At the same time, only 2.67% compounds from the NIH HIV-1 integrase library, which was used as the negative control, have EIIP (0.079 – 0.099 Ry) and AQVN (2.42 - 2.63). Table **3** gives EIIP and AQVN values for CCR5 inhibitors which are in clinical trials. Four of five of these compounds fit the EIIP/AQVN criterion. The only exception is APLAVIROC which satisfies this criterion in terms of EIIP but whose AQVN value of 2.659 is outside of the proposed range.

Table 2. Fraction of Compounds in Different Molecular Libraries which Satisfies the AQVN/EIIP Criterion for CCR5 Inhibitors

Molecular Database	Total number of compounds in library	Number of compounds in filtered area	Compounds in filtered area [%]
Library of CCR5 inhibitors (HIV/OI Therapeutic Database)	884	580	65.61
Library of HIV-1 integrase inhibitors (HIV/OI Therapeutic Database)	1323	30	2.67
PubChem Substances Database (NCBI)	798793	125139	15.73
ChemBank Small Molecules Bioactive Database	4667	868	18.60

Table 3. The Candidate CCR5 Antagonists in Clinical Trial

CCR5 antagonist	Formula	AQVN	EIIP [Ry]	Phase of trial
SCH-C	$C_{28}H_{37}BrN_4O_3$	2.575	0.0868	Phase 2
APLAVIROC	$C_{33}H_{43}N_3O_6$	2.659	0.0985	Phase 3
TAK-779	$C_{33}H_{39}N_2O_2$	2.540	0.0913	Phase 1
MARAVIROC	$C_{29}H_{41}F_2N_5O$	2.436	0.0964	Phase 2/3
VICRIVIROC	$C_{28}H_{38}F_3N_5O_2$	2.500	0.0946	Phase 3

For further validation of this criterion we performed screening of the large PubChem Substances Database (NCBI) and ChemBank Small Molecules Bioactive Database (NCI, NIH) encompassing natural biologically active substances. Of 798793 compounds in the PubChem database, 673654 (84.33%) do not satisfy the EIIP/AQVN criterion for CCR5 inhibitors (Table **2**). A similar result was obtained for compounds from the ChemBank (Table **2**). These results demonstrated that EIIP/AQVN-based VS of molecular libraries can significantly reduce the number of compounds which will be further subjected to more sophisticated and time consuming "VS by docking" or to "similarity-based VS".

An important advantage of EIIP/AQVN-based VS is to avoid selection of candidate CCR5 inhibitors which despite appropriate structural properties cannot effectively prevent HIV infection. This advantage will be illustrated by the following example. In recent years large series of piperidine-

Table 4. Training Set of Piperidine- and Piperazine-Based CCR5 Antagonists [79]

No.	Compound number[a]	Formula	Z*	EIIP [Ry]	K_i[b] [nM]
1	2	$C_{27}H_{35}ClNO_2$	2.424	0.0962	68
2	6	$C_{28}H_{38}N_2O$	2.406	0.0957	110
3	8	$C_{28}H_{38}N_2OS$	2.457	0.0963	86
4	11	$C_{33}H_40N_2O$	2.474	0.0959	360
5	12	$C_{27}H_{33}BrN_2O_2$	2.523	0.0929	54
6	13	$C_{27}H3_4BrN_2O_2$	2.500	0.0946	190
7	14	$C_{28}H_{36}BrN_2O$	2.426	0.0963	62
8	18	$C_{28}H_{36}BrN_3O_2$	2.514	0.0936	2
9	20	$C_{29}H_{38}BrN_3O_2$	2.493	0.0950	2
10	21	$C_{27}HH_{33}BrN_3O_3$	2,612	0.0808	78
11	27	$C_{26}H_{32}ClBrN_4O_2$	2.576	0.0867	3.4
12	30	$C_{28}H_{36}F_4BrN_3O_2$	2.432	0.0964	3
13	32	$C_{28}H_{37}BrN_4O_2$	2.528	0.0924	1.1
14	38	$C_{28}H_{37}BrN_4O_3$	2.575	0.0868	2.1
15	52	$C_{25}H_{30}J_2BrN_3O$	2.484	0.0954	1300
16	61	$C_{28}H_{36}F_3N_3O$	2.422	0.0962	1
17	63	$C_{27}H_{34}F3N_3O_2$	2.493	0.0950	5
18	71	$C_{26}H_{34}F_3N_5O$	2.493	0.0950	3
19	74	$C_{32}H_{34}F_3N_5O$	2.512	0.0937	12
20	77	$C_{34}H_{41}BrN_4O_2$	2.561	0.0888	0.1
21	78	$C_{290}H_{33}BrN_2O$	2.515	0.0935	590
22	79	$C_{26}H_{31}Br_3N_4O_2$	2.576	0.0887	0.3
23	86	$86C_{27}H_{36}BrN_5O_2$	2.563	0.0885	2.2
24	92	$C_{28}H_{38}BrN_5O_2$	2.540	0.0912	2.1
25	100	$C_{28}H_{38}BrN_5O_2S$	2.587	0.0851	1.3

[a]Numbers assigned to compounds in Ref. [79].
[b]K_i – inhibition of RANTES binding to CCR5 [79].

Table 5. The AQVN and EIIP Values Calculated for the Interchim's Library of Piperidines. The Candidate CCR5 Antagonists are Marked in Bold

No.	Piperidine	AQVN	EIIP [Ry]
1	4-(2-Aminoethyl) piperidine	2.160	0.0601
2	4-Cyclopropylamino-1-methylpiperidine	2.207	0.0702
3	3-(3-Piperidine)propionic acid	2.462	0.0962
4	2,2-Dimethyl-3-piperidin-1-yl-propionaldehyde	2.258	0.0798
5	1-Amino-3-(3-methyl-piperidin-1-yl)-propan-2-ol	2.250	0.0784
6	4-Piperidin-1-ylaniline	2.414	0.0960
7	methyl 1-(cyanomethyl)-4-piperidinecarboxylate	2.667	0.0693
8	1-Acetyl-4-cyclopropylaminopiperidine	2.387	0.0948
9	1-n-Propyl-4-cyclopropylaminopiperidine	2.171	0.0628
10	N-(2-Aminoethyl)-2-oxopiperidine-3-carboxamide	2.643	0.0747
11	2-(2,2,6,6-tetramethyl-4-piperidinyl)-1-ethanol	2.167	0.0616
12	phenyl(1-piperidinyl)methanone	2.552	0.0899
13	4-Amino-5-fluoro-2-piperidinopyrimidine	2.593	0.0842
14	Boc-4-aminopiperidine	2.412	0.0959
15	3-(Piperid-4-yl)indole	2.516	0.0934
16	1-(2-Pyrimidinyl)-4-piperidinecarboxylic acid	2.857	0.0102
17	3-Chloro-4-piperidin-1-yl-phenylamine	2.414	0.0960
18	Succinic acid, 2,6-dimethylpiperidide	2.529	0.0923
19	3-Boc-aminomethyl piperidine	2.378	0.0943
20	4-(2-Boc-aminomethyl) piperidine	2.378	0.0943
21	1-Benzylpiperidine-2-carboxylic acid	2.606	0.0819
22	2,4-Dioxo-1-piperidino-1,2,3,4-tetrahydro-5-pyrimidinecarbonitrile	3.000	0.0439
23	6-(2,6-Dioxopiperidin-1-yl)hexanoic acid	2.727	0.0534
24	3-(2-Boc-aminoethyl) piperidine	2.350	0.0921
25	4-(2-Boc-aminoethyl)-1-piperidine	2.350	0.0921
26	1-Benzyl-4-cyclopropylaminopiperidine	2.359	0.0928
27	5,5-Dimethyl-3-[(piperidin-4-ylmethyl)amino]cyclohex-2-en-1-one	2.342	0.0912
28	1-(6-Chloro-3-pyridazinyl)-4-piperidinecarboxamide	2.768	0.0439
29	1-(6-Chloro-3-pyridazinyl)-4-piperidinecarboxylic acid	2.857	0.0102
30	3-(2,6-Dioxopiperidin-1-yl)-4-methylbenzoic acid	3.032	0.0560
31	1-(4-Acetylphenyl)-4-piperidinecarboxylic acid	2.743	0.0488
32	3-(Piperidin-1-ylcarbonyl)bicyclo[2.2.1]hept-5-ene-2-carboxylic acid	2.649	0.0734
33	1-[2,6-Dinitro-4-(trifluoromethyl)phenyl]-4-piperidinecarboxylic acid	3.000	0.0439
34	Methyl 1-(6-chloro-3-pyridazinyl)-4-piperidinecarboxylate	2.774	0.0390
35	1-(6-Chloro-3-pyridazinyl)-4-piperidinecarbohydrazide	2.774	0.0390
36	4-Oxo-4-[(2,2,6,6-tetramethylpiperidin-4-yl)amino]butanoic acid	2.476	0.0958
37	3-(2-Isopropoxycarbonyl-ethyl)-6-oxo-piperidine-3-carboxylic acid	2.757	0.0445
38	1-[(benzyloxy)carbonyl]-4-piperidinecarboxylic acid	2.833	0.0188

(Table 5). contd.....

No.	Piperidine	AQVN	EIIP [Ry]
39	Methyl 1-(4-nitrophenyl)-4-piperidinecarboxylate	2.914	0.0112
40	1-(phenylsulfonyl)-4-piperidinecarboxylic acid	2.700	0.0324
41	1-(Phenylsulfonyl) piperidine-3-carboxylic acid	2.700	0.0324
42	5-Oxo-5-[(2,2,6,6-tetramethylpiperidin-4-yl)amino]pentanoic acid	2.444	0.0964
43	Methyl 1-[4-cyano-5-(methylamino)-3-isothiazolyl]-4-piperidinecarboxylate	2.914	0.0112
44	N-Cyclopropyl-N-(4-piperidinyl)benzenesulfonamide	2.667	0.0693
45	N-Allyl-1-(6-chloro-3-pyridazinyl)-4-piperidinecarboxamide	2.667	0.0693
46	Methyl 1-[1-amino-2-cyano-3-(methylamino)-3-thioxo-1-propenyl]-4-piperidinecarboxylate	2.811	0.0267
47	methyl 1-(phenylsulfonyl)-4-piperidinecarboxylate	2.889	0.0016
48	1-[(4-methylphenyl)sulfonyl]-4-piperidinecarboxylic acid	2.889	0.0016
49	4-(2-Methyl-piperidine-1-sulfonyl)-benzoic acid	2.889	0.0016
50	1-(phenylsulfonyl)-4-piperidinecarbohydrazide	2.889	0.0016
51	1-[(4-fluorophenyl)sulfonyl]-4-piperidinecarboxylic acid	2.970	0.0324
52	Ethyl 2-amino-4-methyl-5-(piperidin-1-ylcarbonyl) thiophene-3-carboxylate	2.750	0.0466
53	methyl 1-[(4-methylphenyl)sulfonyl]-4-piperidinecarboxylate	2.820	0.0233
54	1-[(4-methylphenyl)sulfonyl]-4-piperidinecarbohydrazide	2.820	0.0233
55	1-[(4-methoxyphenyl)sulfonyl]-4-piperidinecarboxylic acid	2.973	0.0337
56	[4-(Piperidine-1-sulfonyl)-phenoxy]-acetic acid	2.973	0.0337
57	1-(3,4-dichlorobenzyl)-4-piperidinecarbohydrazide	2.556	0.0894
58	N-(4-Piperidyl)-3-trifluoromethylsulfonylaniline	2.686	0.0647
59	1-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-piperidinecarboxylic acid	2.688	0.0642
60	methyl 1-[(4-methoxyphenyl)sulfonyl]-4-piperidinecarboxylate	2.900	0.0958
61	1-[(4-nitrophenyl)sulfonyl]-4-piperidinecarboxylic acid	3.257	0.1216
62	4-Chloro-N-cyclopropyl-N-(4-piperidinyl)benzenesulfonamide	2.667	0.0693
63	1-[(4-chlorophenyl)sulfonyl]-4-piperidinecarbohydrazide	2.889	0.0016
64	Methyl 1-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-piperidinecarboxylate	2.629	0.0776
65	1-[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]-4-piperidinecarbohydrazide	2.629	0.0776
66	methyl 1-{[4-(acetylamino)phenyl]sulfonyl}-4-piperidinecarboxylate	2.930	0.0173
67	N-[4-(Piperidine-1-sulfonyl)-phenyl]-succinamic acid	2.930	0.0173
68	Ethyl 5-(4-nitrophenyl)-2-piperidino-1H-pyrrole-3-carboxylate	2.870	0.0056
69	N-Cyclopropyl-N-(4-piperidinyl)-3-(trifluoromethyl)benzenesulfonamide	2.619	0.0795
70	1-[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]-N-(2-hydroxyethyl)-4-piperidinecarboxamide	2.600	0.0829
71	1-[2,6-Dinitro-4-(trifluoromethyl)phenyl]-4-piperidinecarboxylic acid	3.189	0.1064
72	1-[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]-N-(2-hydroxyethyl)-N-methyl-4-piperidinecarboxamide	2.558	0.0891
73	Methyl 1-[2,6-dinitro-4-(trifluoromethyl)phenyl]-4-piperidinecarboxylate	3.100	0.0798

and piperazine-based CCR5 antagonists as HIV-1 entry inhibitors were reported (for review se ref. 81 and references therein). In order to develop predictive pharmacophore models for CCR5 antagonists Debnath selected a training set of 25 piperidine- and piperazine-based compounds (Table **4**) [81]. Based on this learning set, Denath developed a predictive pharmacophore model as a 3D query tool in VS to select new chemical entities as potent CCR5 antagonists. As

Table 6. Data for gp120 Molecules from R5 HIV-1 Isolates [82]

Clade/CRF	Isolate	GenBank Acc No	Coreceptor
A	00KE_KER2018	AF457057	R5
A	99KE_KNH1088	AF457063	R5
A	99KE_KNH1135	AF457065	R5
A	00KE_KNH1144	AF457066	R5
A	00KE_KNH1207	AF457068	R5
A	00KE_KNH1209	AF457069	R5
A	00KE_KSM4030	AF457079	R5
B	94US_33931N	AY713410	R5
B	90US_873	AY713412	R5
B	85US_Ba-L	AY713409	R5
B	92FR_BXO8	AY713411	R5
B	96TH_NP1538	AY713408	R5
B	91US_1	AY173952	R5
B	91US_4	AY173955	R5
C	94IN_20635-4	AY713414	R5
C	98US_MSC5016	AY444801	R5
C	93MW_965	AY713413	R5
C	90SE_364	AY713416	R5
C	89SM_145	AY713415	R5
C	02ET_14	AY255825	R5
C	02ET_288	AY713417	R5
C	01TZ_911	AY253322	R5
C	00TZ_A125	AY253304	R5
C	00TZ_A246	AY253308	R5
D	98UG_57128	AF484502	R5
D	99UG_A03349M1	AF484518	R5
D	99UG_A07412M1	AF484477	R5
D	99UG_A08483M1	AY304496	R5
D	00UG_J32228M4	AF484516	R5
D	00UG_D26830M4	AF484486	R5
D	00UG_E08364M4	AF484487	R5
D	00UG_E13613M4	AF484515	R5
D	00KE_NKU3006	AF457090	R5
CRF01_AE	90TH_CM235	AF259955	R5
CRF01_AE	90TH_CM240	U54771	R5
CRF01_AE	90TH_CM244	AY713425	R5
CRF01_AE	96TH_NI1046	AY713421	R5
CRF01_AE	96TH_NI1149	AY713426	R5
CRF01_AE	98TH_NP1251	AY713422	R5
CRF01_AE	97TH_NP1525	AY713420	Dual
CRF02_AG	98US_MSC5007	AY444808	R5
CRF02_AG	01CM_0002BBY	AY371122	R5
CRF02_AG	01CM_0005BBY	AY371123	R5
CRF02_AG	01CM_0008BBY	AY371124	R5
CRF02_AG	02CM_0013BBY	AY371125	R5
CRF02_AG	02CM_0014BBY	AY371126	R5
CRF02_AG	02CM_0015BBY	AY371127	R5
CRF02_AG	01CM_1475MV	AY371138	R5
CRF02_AG	02CM_1970LE	AY371128	R5
CRF02_AG	91DJ_263	AF063223	R5

can be seen, 24 CCR5 antagonists from this training set fit our EIIP/AQVN criterion. Table **5** gives 73 compounds from the Interchim's piperidine database [www.interchim.com/interchim/chemical/combinatorial_chemistry/hetero/piperidines1.htm]. The EIIP/AQVN-based VS of this database revealed that only 11 (15%) compounds from this list can be selected as candidate CCR5 antagonists. For comparison, the NIH library of CCR5 inhibitors (Table **2**) contains 675 piperidines, and 473 (70%) of these compounds have EIIP (0.079 – 0.099 Ry) and AQVN (2.42 - 2.63).

Considered together, the results presented above strongly suggest that the EIIP/AQVN-based VS, in combination with docking-base VS or similarity-based VS represent a powerful tool for selection of the new CCR5 antagonist for inhibition of HIV-1 entry.

The recently developed combinatorial methods greatly increase the numbers of compounds synthesized and tested but generate very large amounts of data. The average number of compounds synthesized in order to obtain a commercial candidate has risen from 10 000 to around 30-50 000. For this reason, it has become very important to find new methods for extracting useful molecular design information from these large quantities of structure-activity data. For analysis the data sets derived from chemistry and high throughput screening, the quantitative structure-activity relationship (QSAR) is the method of choice, playing a key role in lead optimization.

On the other hand, deriving of QSAR models that are accurate, reliable, and easily interpretable is a difficult task because it is necessary to establish an appropriate balance between model accuracy and complexity. The EIIP/ISM method described here represents a complementary not a concurrent approach to the QSAR methods. Because of its efficacy (by ordinary PC computer about 50000 compounds can be screened in 5 to 10 minutes) this method allows simple and fast prescreening of large molecular libraries and selection of subsets of compounds which will be later treated with more complex and time consuming QSAR methods which use many different descriptors, functional forms, and methods from simple linear equations through to multilayer neural nets.

Selection of Peptide CCR5 Inhibitors Based on the Informational Spectrum Method

In spite of high variability the HIV-1 envelope glycoprotein 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 coreceptor CCR5, we subjected to ISM analysis the gp120 molecules from 50 R5 HIV-1 isolates belonging to different clades and collected from different geographical locations (Table **6**) [82]. In Fig. (**1**) the consensus informational spectrum (CIS) of gp120 molecules from Table **6** is presented. As can be seen, this CIS contains only one characteristic peak corresponding to the IS frequency F(0.1035). This peak frequency represents common biological information encoded by primary structures of analyzed gp120 molecules which probably determines their interaction with the CCR5 coreceptor. In order to prove this assumption, we multiplied the CIS of gp120 with the informational spectra (IS) of CCR5, CXCR4 and CD4 receptors. As can be seen from results of this cross-spectral analysis presented in Fig. (**2**), the characteristic peak corresponding to the frequency F(0.1035) significantly increases after multiplication of CIS of gp120 with IS of CCR5 and decreases after multiplication with IS of CXCR4 and CD4. This result demonstrates that only CCR5 shares common information with CIS of analyzed gp120 molecules, and singles 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 – 186 is essential for information represented by the IS frequency F(0.1035) (Fig. (**3**)). According to previously reported results [37,63], this region of CCR5 is directly involved in interaction with HIV-1 gp120. This conclusion is in accordance with recently reported results of Huang and co-workers demonstrating that the N-terminus of ECL2 is directly involved in binding to V3 loop of HIV-1 gp120 [83]. It is also important to note that most of the CCR5 antagonists which effectively block HIV-1 entry (Table **3**) bind the pocket within ECL2 encompassing residues 168-186. It is also demonstrated that the peptide immunogen derived from ECL2 (residues 169-173) elicits antibodies which efficiently block gp120/CCR5 binding [84,85].

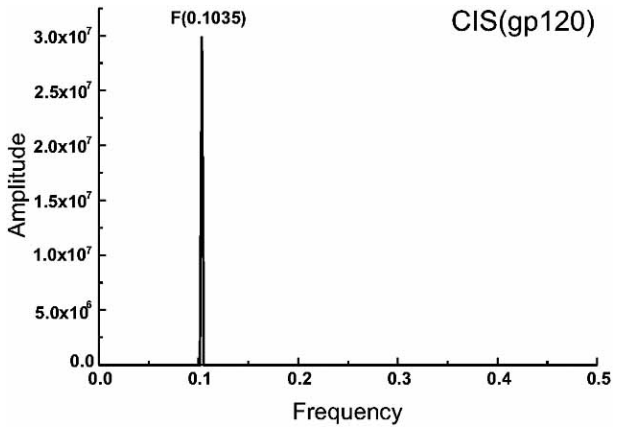

Fig. (1). The consensus informational spectrum (CIS) of 50 R5 HIV-1 gp120 (Table **6**) [80]. The abscissa represents the frequencies from the Fourier transform of the sequence of the electron-ion interaction potential (EIIP). The lowest frequency is 0.00 and the highest is 0.50. The ordinate represents amplitudes, in arbitrary units, corresponding to each frequency component in the informational spectrum (IS).

Screening of Phage Library with ECL2-Derived Peptide

We further used above results of ISM analysis of CCR5 for selected the peptide HIV-1 entry inhibitors. For this purpose we performed the phage display screening using as a target the CCR5-derived peptide RSQKEGLHYT SSSHFPYSQ (residues 168-186). In a process called biopanning, the target is immobilized, in this case *via* an antibody to a tag attached to magnetic beads. After incubation with the phage libraries, phages binding specifically to the target (green) and unspecific binders (red)

are eluted by low pH. Eluted phages are subjected to a negative selection with the same setting except that the target is not present. In this case, target-specific phages remain in the supernatant, whereas unspecific binders are bound. Phages in the supernatant are amplified in E.coli, tested by ELISA for specificity and the phage DNA encoding the peptide inserts is being sequenced for the positive phage clones. Selected phages are incubated with the targets used in positive and negative selections of the biopannings coated on ELISA plates. Binding phages are detected with an anti-M13 monoclonal antibody coupled to horseradish peroxidase for colorimetric detection upon addition of substrate. Target-specific phages will give a positive reaction only in the presence of the target.

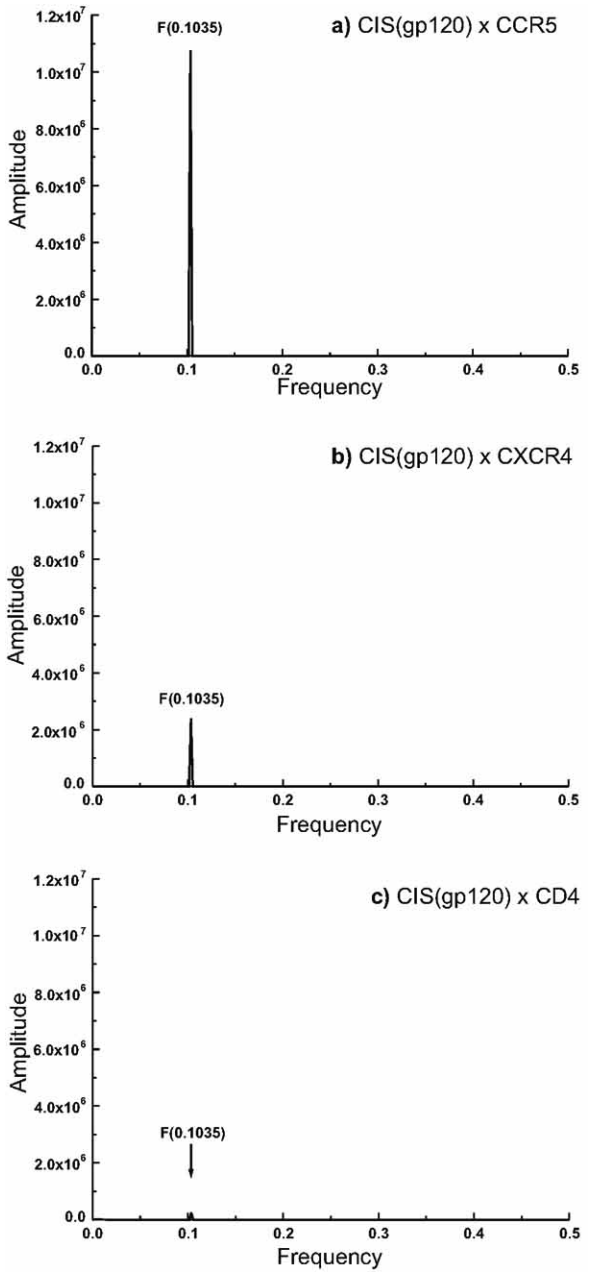

Fig. (2). The cross-spectrum between CIS of R5 HIV-1 gp120 and (a) IS of CCR5, (b) IS of CXCR4 and (c) IS of CD receptor. The abscissa and ordinate are as described in Fig. (**1**).

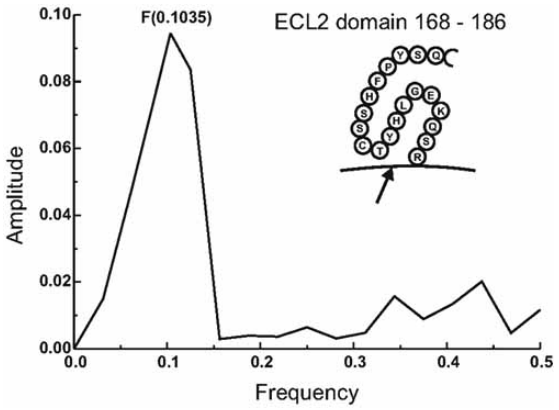

Fig. (3). The informational spectrum of the second extracellular loop (ECL2) of CCR5 encompassing residues 168 – 186. The abscissa and ordinate are as described in Fig. (**1**).

Using the above procedure we selected five phage peptides denoted P13, P17, P21, P25 and P29. As can be seen from CIS of these peptides presented in Fig. (**4**), it contains only one peak corresponding to the frequency F(0.1035). This result points out that information which is essential for interaction between phage displayed peptides and the target ECL2-derived peptide is the same as information which corresponds to CCR5/gp120 interaction.

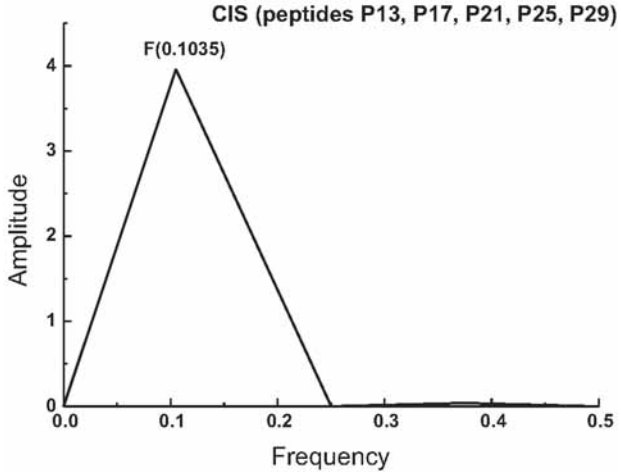

Fig. (4). CIS of five peptides selected by screening of the phage library with ECL2-derived peptide 168-186. The abscissa and ordinate are as described in Fig. (**1**).

Testing of Inhibition of HIV-1 Entry by ECL2 Peptide Ligands

The antiviral activity of the selected phage displayed peptides was analyzed in single-round infection assays using recombinant luciferase reporter viruses pseudotyped with the R5-tropic HIV Env glycoprotein JR-FL. Lentiviral vector supernatants were generated using a three-plasmid transfection system by cotransfecting 293T cells with 7.5 µg of pRRLsinCMV LUC transfer vector, 12.5 µg of the packaging construct containing Gag-Pro-Pol from HIV-1, Tat, Rev and RRE controlled by a CMV promoter and 2 µg of pSVIIl-JR-FL env plasmid (NIH) by the calcium phosphate DNA precipitation method. Cell culture supernatants (DMEM, 2% Glu, 1% Pen/Strep, 5% FCS)

were collected 24h, 36h, 48h and 60h after transfection, cleared through 0.22 µm Millex® HA filter units (Millipore) and concentrated up to 6 times using Vivaspin100 filter tubes and centrifugating 45 min at 2500 rpm at 4°C. Virus particle containing supernatants were titered on U87MG cells (HTB 14, ATCC, Rockville MD) expressing CD4 and CCR5 or CXCR4. The cells were infected for 48h with serial dilutions of the particle stocks in DMEM medium. Infection was quantified by luciferase assay. Cells were washed with PBS, lysed with harvest buffer (0.5 M Mes-Tris, 1M DTT, 10% Triton X-100 and Glycerol) and light emission was measured on a Lumistar Galaxy Luminometer (BMG Labbiotechnologies, Offenburg). Viral stocks were stored at -80°C. Ten to the fourth U87MG-CD4-CCR5 cells were plated per well and incubated for 24h at 37°C/5% CO2. Peptide dilutions were prepared in complete DMEM medium, and were added to the cells at a volume of 20 µl per well. Twenty microliters per well of viral pseudotyped particle dilutions were immediately added to the cells. Cells were incubated for 48h at 37°C/5% CO2 in a final volume of 40 µl of DMEM medium containing the peptide inhibitors and the pseudotyped viral particles. Cells were then lysed for determination of luciferase activity using beetle luciferin (Promega) in luciferase buffer (0.5 M Mes-Tris, 0.5 M MgCl2, ATP). Inhibition was expressed as the percentage of the positive control (containing virus but no peptide) after subtraction of the negative control without virus.

The best HIV-1 entry inhibition is obtained with peptide P25. According to results presented in Fig. (**5**), in the nanomolar range this peptide is even more effective than peptide T-20 (FUZEON) representing the first commercially available HIV-1 entry inhibitor.

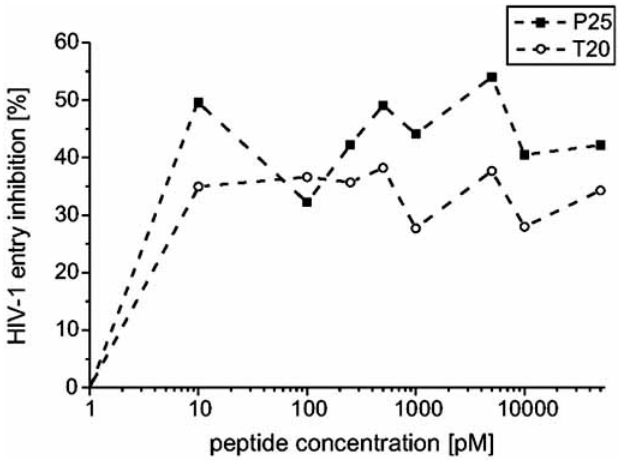

Fig. (5). Entry inhibition of R5 HIV-1 isolate JR-FL with peptide P25 selected by screening of the phage library and the commercial HIV-1 entry inhibitor peptide T-20.

Taking together, these results point out the ISM approach as a powerful tool for selection of peptide HIV-1 entry inhibitors.

CONCLUSION

A corollary of results presented in this review is that EIIP/AQVN and ISM approaches can be effectively used together with other complementary structural approaches for selection of candidate HIV-1 entry inhibitors.

ACKNOWLEDGMENTS

We are grateful to Dr. Hans.Langedijk from Pepscan, System B.V. for synthesis of peptides for HIV inhibition experiments and to Dr. Nils Gunnar Johansson from Medivir for encouragement and critical advice. This work was supported in part by European Commission project TRIoH LSHG-CT-2003-503480 and Serbian Ministry of Science grant 143001g.

REFERENCES

[1] Dimasi, J.A. *Clin. Pharmacol. Ther.*, **2001**, *69*, 297.
[2] Dimasi, J.A.; Hansen, R.W.; Grabowski H.G. *J. Health Econ.*, **2003**, *22*, 151.
[3] Augen, J. *Drug Discov.Today*, **2002**, *7*, 315.
[4] McGee, P. *Drug Discov.Today*, **2005**, *8*, 23.
[5] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. *J. Mol. Biol.*, **1982**, *161*, 269.
[6] Goodsell, D.S.; Morris, G.M.; Olson, A.J. *J. Mol. Recognit.*, **1996**, *9*,1.
[7] Burkhard, P.; Taylor, P.; Walkinshaw, M.D. *J. Mol. Biol.*, **1998**, *277*, 449.
[8] Lorber, D.M.; Shoichet, B.K. *Protein Sci.*,**1998**, *7*, 938.
[9] Kramer, B.; Rarey, M.; Lengauer, T. *Proteins*, **1999**, *37*, 228.
[10] Taylor, J.S.; Burnett, R.M. *Proteins*, **2000**, *41*, 173.
[11] Ewing, T.J.; Makino, S.; Skillman, A.G.; Kuntz, I.D. *J. Comput. Aided. Mol. Des.*, **2001**, *15*, 411.
[12] Jiang, F.; Kim, S.H. *J. Mol. Biol.*, **1991**, *219*, 79.
[13] Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Biomol. Struct. Dyn.*, **1991**, *8*, 1267.
[14] Dunbrack, R.L.Jr.; Karplus, M. *J. Mol. Biol.*, **1993**, *230*, 543.
[15] Leach, A.R. *J. Mol. Biol.*, **1994**, *235*, 345.
[16] Leach, A.R.; Lemon, A.P. *Proteins*, **1998**, *33*, 227.
[17] Kallblad, P.; Dean, P.M. *J. Mol. Biol.*, **2003**, *326*, 1651.
[18] Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J.D. *Molecular Biology of the Cell*, Garland: New York, **1989**.
[19] Chapeville, F.; Haenni, A.L. (Editors) *Chemical Recognition in Biology*. Springer-Verlag: Berlin, **1980.**
[20] Frohlich, H. *Int. J. Quantum Chem.*, **1968**, *2*, 641.
[21] Frohlich, H. *Nature*, **1970**, *228*, 1093.
[22] Frohlich, H. *Proc. Natl. Acad. Sci.*, **1975**, *72*, 4211.
[23] Moskalenko, S.A.; Miglei, M.F.; Khadshi, P.I.; Pokatilov, E.P.; Kiselyova, E.S. *Phys. Lett.*, **1980**, *76A*, 197.
[24] Veljkovic, V. *A theoretical approach to preselection of carcinogens and chemical carcinogenesis*. Gordon & Breach: New York, **1980**.
[25] Veljkovic, V.; Lalovic, D. *Cancer Biochem. Biophys*., **1976**, *1*, 295.
[26] Veljkovic, V.; Slavic, I. *Phys. Rev. Lett.*, **1972**, *29*, 105.
[27] Veljkovic, V. *Phys. Lett.*, **1973**, *45A*, 41.
[28] Veljkovic, V.; Lalovic, D. *Phys. Lett.*, **1973**, *45A*, 59.
[29] Veljkovic, V.; Lalovic, D. *Cancer Biochem. Biophys.*, **1976**, *1*, 295,
[30] Veljkovic, V.; Lalovic, D. *Experientia*, **1977**, *33*, 1228.
[31] Veljkovic, V.; Lalovic, D. *Experientia*, **1977**, *34*, 1342.
[32] Veljkovic, V.; Ajdacic, V. *Experientia*, **1977**, *34*, 639.
[33] Ajdacic, V.; Veljkovic, V. *Experientia*, **1977**, *34*, 633.
[34] Veljkovic, V.; Cosic, I.; Dimitrijevic, B.; Lalovic, D. *IEEE Trans.*, **1985**, *BME-32* , 337.
[35] Cosic, I.; Nesic, D.; Pavlovic, M.; Williams, R. *Biochem. Biophys. Res. Commun*., **1986**, *141*, 831.
[36] Veljkovic, V.; Cosic, I. *Cancer Biochem. Biophys*., **1987**, *9*, 139.
[37] Veljkovic, V.; Metlas, R. *Cancer Biochem. Biophys.*, **1988**, *10*, 191.
[38] Skerl, V.; Pavlovic, M. *FEBS Lett.*, **1988**, *239*, 141.
[39] Cosic, I.; Nesic, D. *Eur. J. Biochem.*, **1988**, *170*, 247.
[40] Cosic, I.; Pavlovic, M.; Vojisavljevic, V. *Biochemie*, **1989**, *71*, 333.

[41] Veljkovic, V.; Lalovic, D. *Biosystems*, **1990**, *23*, 311.
[42] Cosic, I. In *Bioinstrumentation and Biosensors*, Wise, D. Ed.; Marcel Dekker: New York, **1990**; pp. 475-485.
[43] Cosic, I.; Hodder, A.; Aguilar, M.; Hearn, M. *Eur. J. Biochem*., **1991**, *198*, 113.
[44] Cosic, I.; Hearn, M. *J. Mol. Rec.*, **1991**, *4*, 57.
[45] Veljkovic, V.; Metlas, R.; Raspopovic, J.; Pongor, S. *Biochem. Biophys. Res. Commun*., **1992**, *189*, 705.
[46] Cosic, I. *Trans. Biomed. Eng.*, **1994**, *12*, 1101.
[47] Cosic, I.; Drummond, A.E.; Underwood, J.R.; Hearn, M. *Mol. Cell. Biochem.*, **1994**, *130*, 1.
[48] Cosic, I.; Birch, S. *IEEE EMBS*, **1994**, *16*, 265.
[49] Nair, M.T.; Tambe, S.S.; Kulkarni, B.D. *FEBS Lett.*, **1994**, *346*, 273.
[50] Veljkovic, V.; Johnson, E.; Metlas, R. *Cancer J.*, **1995**, *8*, 308.
[51] Cosic, I. *Biotechnology*, **1995**, *13*, 236.
[52] Birch, S.; West, R.; Cosic, I. *Australas. Phys. Eng. Sci. Med.*, **1995**, *18*, 197.
[53] Lazovic, J. *Comp. Appl. Bio. Sci.*, **1996**, *16*, 553.
[54] Shepherd, A.J.; Gorse, D.; Thornton, J.M. *Proteins*, **2003**, *50*, 290.
[55] This journal abb. is correct, for ref see: http://www.efm.leeds.ac.uk/~mark/ISIabbr/I_abrvjt.html
[56] Pirogova, E.; Simon, G.P.; Cosic, I. *IEEE Trans. Nanobioscience*, **2003**, *2*, 63.
[57] Fang, Q.; Cosic, I. *Australas. Phys. Eng. Sci. Med.*, **1998**, *21*, 179.
[58] de Trad, C.H.; Fang, Q.; Cosic, I. *Biophys. Chem.*, **2000**, *84*, 149.
[59] Parbhane, R.V.; Unniraman, S.; Tambe, S.S.; Kulkarni, B.D. *J. Biomol. Struct. Dyn.*, **2000**, *17*, 665.
[60] Murakami, M. *J. Prot. Chem.*, **2000**, *19*, 609.
[61] de Trad, C.H.; Fang, Q.; Cosic, I. *Protein Eng.*, **2002**, *15*, 193.
[62] Arrigo, P.; Ivaldi, G.; Cardo, P.P. *In Silico Biol*., **2002**, *2*, 143.
[63] Veljkovic, N.; Branch,D.R.; Metlas, R.; Prljic, J.; Vlahovicek, K.; Pongor, S.;Veljkovic, V. *J. Pept. Res.*, **2003**, *62*, 1.
[64] Shepherd, A.J.; Gorse, D.; Thornton, J.M. *Proteins*, **2003**, *50*, 290.
[65] Chen, J.; Li, H.; Sun, K.; Kim, B. *IEEE Signal. Proc. Mag.*, **2003**, *20*, 106.
[66] Pirogova, E.; Simon, G.P.; Cosic, I. *IEEE Trans Nanobioscience*, **2003**, *2*, 63.
[67] Ivanovic, V.; Demajo, M. Todorovic-Rakovic, N.; Nikolic-Vukosavljevic, D.; Neskovic-Konstantinovic, Z.; Krtolica, K.; Veljkovic, V.; Prljic, J.; Dimitrijevic B. *Med. Hypoth*., **2004**, *62*, 727.
[68] Vaidyanathan, P.P.; Yoon, B. *J. Franklin Inst*., **2004**, *341*, 111.
[69] Sauren, M.; Pirogova, E.; Cosic, I. *Austral. Phys. Eng. Sci. Med.*, **2004**, *27*, 174.
[70] Krishnan, A.; Li, K.B.; Issac, P. *In Silico Biol.*, **2004**, *4*, 13.
[71] Vaidyanathan, P.P. *IEEE Circuits Syst. Mag.*, **2004**, *4*, 6.
[72] Wen, Z.; Wang, K.; Li, F.; Nie, F.; Yang, Y. *Comput. Biol. Chem.*, **2005**, *29*, 220.
[73] Guo, Y.; Li, M.; Wang, K.; Wen, Z.; Lu, M.; Liu, L.; Jiang, L. *Acta Biochim. Biophys. Sin. (Shanghai)*, **2005**, *37*, 759.
[74] Rakovic, D.; Dugic, M.; Plavsic, M. *Mater. Sci. Forum*, **2005**, *494*, 513.
[75] Riaz, T.; Li, K.B.; Tang, F.; Krishnan, A. *In Silico Biol.*, **2005**, *5*, 3.
[76] Huang, J.; Tan, P.H.; Li K.B.; Matsumoto, K.; Tsujimoto, M.; Bay, B.H. *Int. J. Oncol.*, **2005**, *26*, 607.
[77] Veljkovic, N.; Perovic, V. *Prot. Pept. Lett.*, **2006**, *13*, 807.
[78] Seifert, M.H.J.; Wolf, K.; Vitt, D. *Biosilico*, **2003**, *1*, 143.
[79] Berger, E.A.; Murphy, P.M.; Farber, J.M. *Annu. Rev. Immunol.*, **1999**, *17*, 657.
[80] Li, H.; Sutter, J.; Hoffman, R. In *Pharmacophore Perception, Development and Use in Drug Design.*, Guner, O.F. Ed.; International University Line: La Jolla, CA, **1999**; pp 137-189.
[81] Debnath, A.K. *J. Med. Chem.*, **2003**, *46*, 4501.
[82] Brown, B.K.; Darden, J.M.; Tovanabutra, S.; Oblander, T.; Frost, J.; Sanders-Buell, E.; de Souza, M.S.; Birx, D.L.; McCutchan, F.E.; Polonis, V.R. *J. Virol*., **2005**, *79*, 6089.
[83] Huang, C.C.; Tang, M.; Zhang, M.Y.; Majeed, S.; Montabana, E.; Stanfield, R.L.; Dimitrov, D.S.; Korber, B.; Sodroski, J.; Wilson, I.A.; Wyatt, R.; Kwong, P.D. *Science*, **2005**, *310*, 1025.
[84] Misumi, S.; Endo, M.; Mukai, R.; Tachibana, K.; Umeda, M.; Honda, T.; Takamune, N.; Shoji, S. *J. Biol. Chem.*, **2003**, *278*, 32335.
[85] Nakayama, D.; Misumi, S.; Mukai, R.; Tachibana, K.; Umeda, M.; Shibata, H.; Takamune, N.; Shoji, S. *J. Biochem.*, **2005**, *138*, 571

Received: August 01, 2006 Revised: November 10, 2006 Accepted: November 11, 2006