SEQUENCE UNIQUENESS AS A MOLECULAR SIGNATURE OF HIV-1-DERIVED B-CELL EPITOPES

A. LUCCHESE, R. SERPICO, V. CRINCOLI[1], Y. SHOENFELD[2] and D. KANDUC[3]

Department of Odontostomatology, Orthodontics and Surgical Disciplines, University of Naples (SUN), Italy; [1]Department of Odontostomatology and Surgery, University of Bari, Italy; [2]Center for Autoimmune Diseases, Department of Medicine 'B', Sheba Medical Center, Israel and Sackler Faculty of Medicine, Tel-Aviv University, Israel; [3]Department of Biochemistry and Molecular Biology, University of Bari, Italy

Received December 18, 2008 – Accepted June 18, 2009

The complex pathophysiology of human immunodeficiency virus (HIV) infection and the relatively high mutation rate of the retrovirus make it challenging to design effective anti-HIV vaccines. Several attempts have been made during the last decades to elucidate the enigmatic immunology of HIV infection and to predict potential immunogenic peptides for active vaccination using bioinformatic analysis methods. The results obtained to date to address this important problem are scarce. In this study, we exploit available HIV databases and analyse previously characterized HIV-encoded linear B-cell epitopes for their amino acid sequence similarity to the human or murine host proteome. We obtained further documentation that the HIV-derived antibody-targeted sequences mostly coincide with peptide areas rarely shared with the host proteins. *In toto*, our past and present data give clear-cut support to the statement that low-similarity to the host proteome is a major mechanism in defining viral peptide immunogenicity and indicate a possible way for inducing effective, high-titer, and non-cross-reactive antibodies to be used in anti-HIV vaccine therapy.

HIV infection is characterized by an initial humoral immune response that can contribute to reducing viral load but does not produce viral clearance (1-3). This suggests that a more vigorous antibody (Ab) response might presumably lead to HIV eradication/control. Indeed, broad neutralizing anti-HIV Abs have been shown to suppress immune deficiency virus infection in macaques (4) and, at least transiently, in humans (5).

In this context, understanding the features of the anti-HIV humoral immune response is of particular interest. As a continuation of our studies on the identification of HIV B-cell epitopes (6), we carried out a thorough analysis of the already well-characterized HIV-1-derived B-cell epitopic sequences by exploring the sharing of amino acid motifs between human and viral proteins. We report that almost all of the linear determinants recognized by mono- and/or polyclonal anti-HIV-1 Abs harbor pentapeptide motifs with no or very limited sequence similarity to the host proteome, thus indicating that a potential peptide immunogenicity is inversely

Key words: HIV-1 B-cell epitopes, similarity level, sequence uniqueness, proteome scanning

Mailing address: Dr D. Kanduc,
Department of Biochemistry
and Molecular Biology,
University of Bari, Via Orabona 4,
70126 Bari, Italy
Tel: ++39 0805443321 Fax: ++39 0805443317
e-mail: d.kanduc@biologia.uniba.it

0394-6320 (2009)

correlated with its incidence in the proteome.

MATERIALS AND METHODS

Similarity analyses to the human (or murine) proteome were conducted on HIV-1 B-cell epitopes catalogued in the www.hiv.lanl.gov/content/immunology website (7). HIV-derived peptidic epitopes were dissected into pentamer motifs that were analyzed *versus* the human (or murine) proteome for exact identical pentamer matches using PIR protein database and perfect peptide match program (pir.georgetown.edu) (8). Viral pentamers sequentially overlapped by four residues, i.e., ERYLK, RYLKD, YLKDQ, etc.

We used pentamer motifs as probes for proteome scanning since the scientific literature indicates 5 to 6 amino acids as sufficient minimal antigenic determinants critically involved in immune recognition (9-10). The similarity level of each viral pentamer sequence to a proteome was quantified as the number of times the pentamer occurs in the analysed proteome. More precisely, the similarity level of a viral 5-mer is considered zero when the viral 5-mer is absent in the proteome under analysis, whereas the similarity level of a viral 5-mer is high when the viral 5-mer is repeatedly represented in the protein set that comprehensively forms the proteome under analysis. As a rule, the proteomic similarity profile of a protein primary sequence presents a fractal behaviour with zero/low-similarity regions that alternate to high-similarity areas with even hundreds of perfect matches (11). A pentapeptide fragment having up to 5 total perfect matches to the host proteome was considered as a low similarity sequence (12-15).

RESULTS

Fig. 1 shows the similarity profile to the mouse proteome of the HIV-1 protease$_{21\text{-}53}$EALLDTGAD DTVLEEMSLPGRWKPKMIGGIGGF sequence.

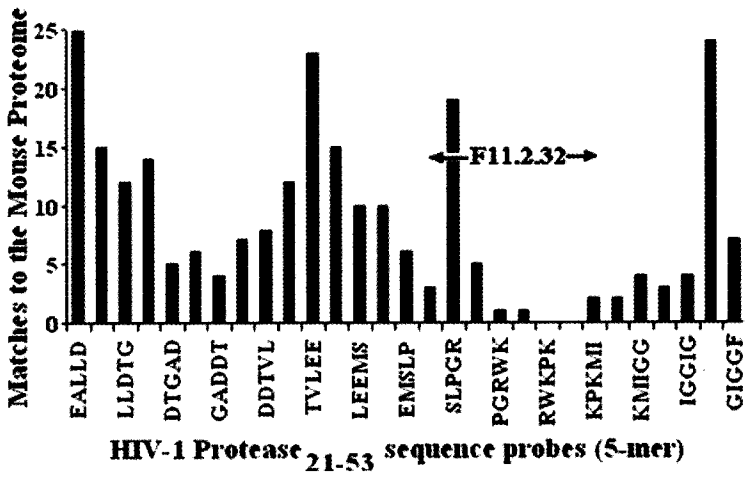

Fig. 1. *Similarity profile to the murine proteome of the HIV-1 protease$_{21\text{-}53}$EALLDT GADDTVLEEMSLPGRWKPKMIGGIGGF sequence and epitope mapping of the mouse mAb F11.2.32. The epitope is indicated by a double arrowed line, with the line length indicating the epitopic length and the arrows illustrating the epitope location along the viral sequence. The MSLPGRWKPKM epitopic area targeted by the anti-HIV-1 protease mAb F11.2.32 coincides with a peptide area endowed with low-similarity to the murine proteome.*

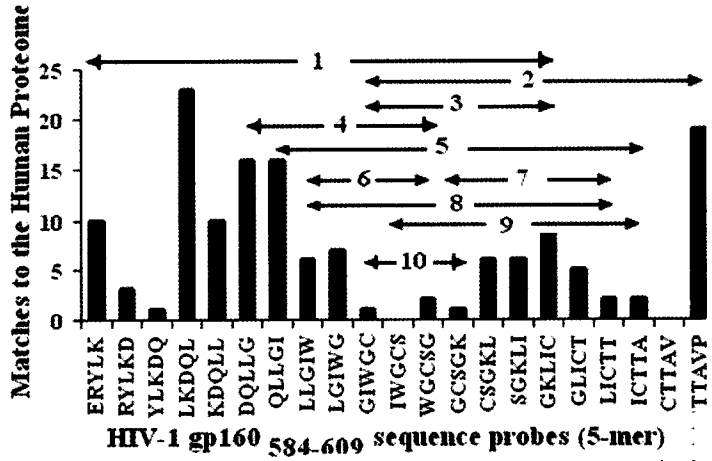

Fig. 2. *Similarity profile to the human proteome of the immunogenic HIV-1 gp160$_{584\text{-}609}$ERYLK DQLLGIWGCSGKLICTTAVP sequence and epitope mapping of 10 human Ab preparations. For each epitope, location and length along the HIV-1 gp160$_{584\text{-}609}$ sequence are represented by a double-arrowed line. The line length indicates the epitopic length and the arrows illustrate the exact epitope location along the viral sequence. Human Ab preparations are numbered from 1 to 10*

Table I. *Epitope mapping of 10 human Ab preparations along the HIV-1 gp160$_{584-609}$ ERYLKDQLLGIWGCSGKLICTTA VP sequence, as a function of amino acid sequence similar to the human proteome.*

Human Ab[a]: N.	Designation:	AA Epitope Sequence[b,c]	Matches to the Human Proteome
1	polyclonal	erylkdqllgIWGCSgklic	0
2	polyclonal	gIWGCSgklicttavp	0
3	1B8.env	gIWGCSgklic	0
4	F11	dqllgIWGCSg	0
5	polyclonal	qllgIWGCSgklictta	0
6	240-D	llgIWGCSg	0
7	clone 3	GCSGKlictt	1
8	F240	llgIWGCSgklictt	0
9	polyclonal	IWGCSgklictta	0
10	α598-609	gIWGCSgk	0

[a] *Ab numbering as reported in Fig. 1.*

[b] *The lowest similarity pentapeptide in each epitope is reported and given in capital letters.*

[c] *The sequence alignment indicates the epitopic commonality of low-similarity fragment(s).*

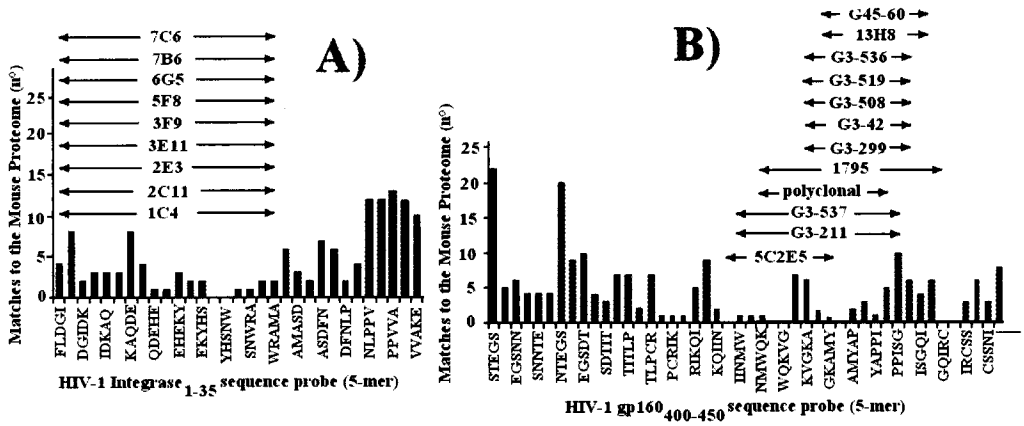

Fig. 3. *Epitope targeting by murine Abs along HIV-1 polyprotein sequences as a function of amino acid sequence similarity to the mouse proteome. HIV-1 antigen sequence: **A)** integrase, aa 1-35; **B)** gp160, aa 400-450. Each epitope is represented by a double-arrowed line reporting the Ab designation. The line length indicates the epitope length and the arrows illustrate the exact epitope location along the viral sequence.*

In Fig. 1 the double-arrowed line represents the length and location of the murine mAb F11.2.32 epitope and indicates that the HIV-1 protease$_{21-53}$ sequence is targeted by the murine mAb at level of the MSLPGRWKPKM fragment (7, 16-17). It can be seen from Fig. 1 that the HIV-1 protease$_{36-46}$MSLPGRWKPKM epitope represents a low similarity area in the HIV-1 protease$_{21-53}$EALLDTGA DDTVLEEMSLPGRWKPKMIGGIGGF sequence. It is worth noting the MSLPGRWKPKM epitope contains two pentamers, RWKPK and WKPKM, that have no match at all with the mouse proteins, thus representing unique pentapeptide signatures of the viral sequence.

The similarity analysis described in Fig. 1 for HIV-1 protease$_{36-46}$MSLPGRWKPKM epitope was extended to a number of HIV-1 epitopes reported in the hiv.lanl.gov/content/immunology website.

Table II. *Similarity level to the host proteome of HIV-1 Ab binding sites.*

Antibody	Species[a]	HIV-1 Protein	Epitopic Sequence[b,c]	Matches to the Host Proteome
3B10	M	P17	irlrpggkkkyMLKHVvwaa	0
11H9	M	p17	lekieeEQNKSkkka	2
1D9	M	p17	aagtgnssqVSQNY	3
F5-2	M	p24	aisprTLNAW	4
15F8C7	M	p24	atpqDLNTMl	2
111/073	M	p24	etineeAAEWD	0
13-102-100	M	p24	HPVHAgpiapg	2
38:9.6K	M	p24	npPIPVGeiy	2
23A5G4	M	p24	irqgpkepfrdyvdRFYKTl	0
11C10B10	M	p24	tlraeqasqEVKNWm	0
BE3	M	p24	qasqevKNWMTetll	0
LH-104-B	M	p24	GHKARv	2
1.153 G10	M	RT	KTGKY	4
RTMAb8	M	RT	ttESIVIw	1
1C4	M	Integrase	fldgidkaqDEHEKyh	1
17	M	Integrase	DFNLPpvvake	2
8-22	M	Integrase	gpakllwkgegAVVIQ	2
4-20	M	Integrase	dnsdIKVVP	1
6-19	M	Integrase	rrkAKIIRd	4
TB12	M	Tat	gISYGRkkrrqrrppqg	1
133/290	M	gp160	YDTEVhnvwa	1
D/5E12	M	gp160	aCVPTDpnpqevvlvnvten	0
T7.1	M	gp160	eNFDMWkndm	0
T9	M	gp160	enfdmWKNDM	0
B35	M	gp160	FNMWKn	1
187.2.1	M	gp160	veQMHEDiislwdqslkpcv	0
W1	M	gp160	eqmhediislwdqslKPCVK	0
B33	M	gp160	tplcvslkctdlgnATNTNs	0
697-D	H	gp160	istsirgkvqkEYAFFykld	0
G3-136	M	gp160	qkEYAFFykld	0
493-156	M	gp160	epipiHYCAPagfailkcnn	0
110.E	M	gp160	ngslaeeevvirsvNFTDNa	0
IIIB-V3-21	M	gp160	iNCTRP	1
MO97/V3	H	gp160	PNNNTrksir	2
Polyclonal	H	gp160	nynkrkrihigpgRAFYTtk	0
9284	M	gp160	NNTRKsiriqrg	2
1324-E	H	gp160	TRTSVr	4
MO99/V3	H	gp160	RKSIR	3
2C4	M	gp160	rkrIHIGPgrafytt	0
391/95-D	H	gp160	krihigpGRAFY	0
Polyclonal	H	gp160	rirpgrAFVTIgk	2
1027-15D	H	gp160	ksITKGP	3
M77	H	gp160	iriqrgpgrAFVTI	2
Polyclonal	H	gp160	rihIGPGRafyt	0
loop 2	H	gp160	sisgpgRAFYTg	0
MN215	H	gp160	rihIGPGRafyttkn	0
Nea 930i	M	gp160	riqrgpgraFVTIGki	2
504-D	H	gp160	ihIGPGR	0
83.1	M	gp160	IYIGPgr	2
M5.5	M	gp160	IHIGPgrafyt	0
268-D	H	gp160	hIGPGR	0

Antibody	Species[a]	HIV-1 Protein	Epitopic Sequence[b,c]	Matches to the Host Proteome
polyclonal	M	gp160	gpGRAFY	2
N70-1.9b	H	gp160	pGRAFY	0
polyclonal	M	gp160	nMWQEVgkamya	0
G3-299	M	gp160	evGKAMYapp	1
13H8	M	gp160	gkamYAPPIs	1
Polyclonal	M	gp160	kamyappisGQIRCssnitg	0
Polyclonal	H	gp160	NNNNGsei	4
CRA1	M	gp160	snnESEIFrl	2
9201	M	gp160	gggdMRDNWrse	0
H11	M	gp160	GGDMRd	1
W2	M	gp160	ggdmrDNWRSelykykvvki	0
110.1	M	gp160	IEPLGvaptk	1
1331A	H	gp160	dWVVQRekr	1
Polyclonal	H	gp160	lqarilaverYLKDQql	1
240-D	H	gp160	llgIWGCSg	1
115.8	M	gp160	lglIWGCSgklic	1
Clone 3	H	gp160	GCSGKlictt	1
105-732	M	gp160	kgRLICYt	1
3D6	H	gp160	sgklictTAVPWnas	0
5B2	M	gp160	eLDKWA	2
Polyclonal	M	Nef	sVIGWLtvrermrrae	0
A6	M	Nef	ERMRRaepa	4
Polyclonal	M	Nef	aatnAACAWleaqeee	0
Polyclonal	M	Nef	gyFPDWQnytpgpgv	0
F3	M	Nef	tpgpgVRYPL	1
F1	M	Nef	vEPDKVeean	2
E7	M	Nef	hhvarelHPEYFknc	0
EH1	M	Nef	marelhPEYYKdc	0

[a]*Species: H, human; M, mouse.* [b, c] *See legend to Table 1.*

In agreement with the data of Fig. 1, it was found that the Ab-targeted sequences mostly coincide with peptide areas rarely shared with the host proteome. As a representative example, Fig. 2 illustrates the similarity profile to the human proteome of the highly immunogenic HIV-1 $gp160_{584-609}$ ERYLKD QLLGIWGCSGKLICTTAVP cluster (7, 18-22), and the epitope mapping of the humoral human immune response. In Fig. 2, each epitope is represented by a double-arrowed line, with the line length indicating the epitopic length, and arrows illustrating the exact epitope location along the HIV-1 $gp160_{584-609}$ sequence. It can be seen that the 10 epitopes targeted by the 10 different human Ab preparations are differently dislocated along the $gp160_{584-609}$ sequence and have different amino acid length (from 8 to 20 aa), as reported in further detail in Table I. At the same time, it is also evident that Fig. 2 and Table I clearly illustrate that the 10 different epitopes of the human immune response are unified by the common feature of being all endowed with fragment(s) with low-similarity to the human proteins. More precisely, Fig. 2 and Table I clearly delineate the low-similarity GIWGCSGK region as the core of the immune recognition by the 10 human Ab preparations.

Similar data were obtained when HIV-1 epitopes of mouse Abs were analyzed (Fig. 3) (7, 23-24). Fig. 3, panel A, indicates that 9 different murine Ab preparations targeted the same HIV-1 integrase sequence (aa 1-23) characterized by the presence of zero similarity pentamer fragments (KYHSN and YHSNW). Likewise, Fig. 3, panel B, shows that epitopes of 12 murine Abs are located along a low-similarity HIV-1 gp160 sequence (aa 426-445)

hosting 5-mers with zero similarity to the mouse proteome.

At a general level, when the identity analysis was carried out by examining the humoral immune response for the various HIV-1 proteins in different hosts, the relationship between epitopicity and low-similarity to the host assumes the value of a rule. Indeed, a thorough survey of scientific literature on HIV-1 epitopes (7; hiv.lanl.gov website) produced the data illustrated in Table II, according to which low identity score is a common denominator for HIV-1-derived determinants described in a number of reports. It clearly emerges from Table II that zero/low similarity to the host proteome characterizes HIV-1 B-cell epitopes recognized by anti-HIV mono- and/or polyclonal Abs raised in murine and/or human organisms.

DISCUSSION

The immunogenic potential of an amino acid sequence relies on different properties derived from a set of complex structural interactions and pathways. To cite only a few, physico-chemical modules such as structural protrusion, hydrophobicity and hydrophilicity index, helical pattern, scales of turns, flexibility scales, may characterize/alter immunoreactive sequences (25-28). Epitope immunoreactivity is also biased by the conformation and post-translational modifications of the antigen (folding, phosphorylation, glycosylation, methylation, etc.) (13, 15). Crypticity and proteolysis add further complexity to the potential definition of peptide immunogenicity (13, 15). It has been proposed that B-cell epitopic peptides are characterized by a low sequence similarity to the host proteome (12). In this study we validate previous data indicating that proteomic similarity to the host proteome shapes and modulates the B-cell epitope repertoire in HIV-associated humoral immune response (6), in this way confirming the low-similarity hypothesis (12).

These data are of special importance when analyzed in the context of the social and economic disruption caused by HIV/AIDS pandemic for millions of people. Since the discovery of HIV almost three decades ago, an enormous amount of research and economical resources have been devoted to the search of effective anti-HIV therapeutic/preventive vaccines. It is all the more extraordinary, therefore, that the development of a vaccine has completely failed in the case of HIV (29).

The wide and detailed HIV-1 epitope analysis reported here demonstrates that low similarity to the host proteome is an almost constant characteristic of the most part of HIV-1-derived B-cell epitopes currently catalogued. Therefore, the present results not only further validate the scientific rationale according to which only those sequences within a foreign protein that are not shared with the host proteome might theoretically have the potential to evoke an immune reaction (12), but also offer new immuno-therapeutical approaches in the fight against HIV. Indeed, from the clinical point of view, this study is definitely of relevance for the rational development of safe and effective anti-HIV Ab-based treatments. The zero similarity profiling of viral epitopic portions to the human proteome appears to be the optimal premise for defining an HIV-1 peptidome platform theoretically able to evoke powerful humoral immune response without concomitant harmful side reactions (30).

REFERENCES

1. Weiss RA, Clapham PR, Cheingsong-Popov R, Dalgleish AG, Carne CA, Weller IV, Tedder RS. Neutralization of human T lymphotropic virus type III by sera of AIDS and AIDS-risk patients. Nature 1985; 316:69-72.
2. Moog C, Fleury HJ, Pellegrin I, Kirn A, Aubertin AM. Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. J Virol 1997; 71:3734-41.
3. Pantaleo G, Demarest JF, Schacker T, et al. The qualitative nature of the primary immune response to HIV infection is a prognosticator of disease progression independent of the initial level of plasma viremia. Proc Natl Acad Sci USA 1997; 94:254-8.
4. Emini EA, Schleif WA, Nunberg JH, et al. Prevention of HIV-1 infection in chimpanzees by gp120 V3 domain-specific monoclonal antibody. Nature 1992; 355:728-30.
5. Trkola A, Kuster H, Rusert P, et al. Delay of HIV-1 rebound after cessation of antiretroviral therapy

through passive transfer of human neutralizing antibodies. Nat Med 2005; 11:615-22.
6. Kanduc D, Serpico R, Lucchese A, Shoenfeld Y. Correlating low-similarity peptide sequences and HIV B-cell epitopes. Autoimmun Rev 2008; 7:291-6.
7. Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI. HIV Molecular Immunology 2006/2007. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, 2007; LA-UR 07-4752.
8. Wu CH, Huang H, Arminski L, et al. The Protein Information Resource: an integrated public resource of functional annotation of proteins. Nucleic Acids Res 2002; 30:35-7.
9. Oldstone MB. Molecular mimicry and immune-mediated diseases. FASEB J 1998; 12:1255-65.
10. Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D. Peptidology: short amino acid modules in cell biology and immunology. Amino Acids 2007; 33: 703-7.
11. Kanduc D, Capone GM. The similarity profile of the human proteome as a fractal dimension. Biol Forum 2008; 101:142-5.
12. Kanduc D. Immunogenicity in peptide-immunotherapy: From self/nonself to similar/ dissimilar sequences In Multichain immune recognition receptor signaling: from spatiotemporal organization to human disease A.B. Sigalov, ed., Landes Bioscience, Austin, TX. 2008; p. 198.
13. Mittelman A, Tiwari R, Lucchese G, Willers J, Dummer R, Kanduc D. Identification of monoclonal anti-HMW-MAA antibody linear peptide epitope by proteomic database mining. J Invest Dermat 2004; 123: 670-5.
14. Kanduc D. "Self-nonself" peptides in the design of vaccines. Curr Pharm Des 2009; 15:3283-9.
15. Stufano A, Kanduc D. Proteome-based epitopic peptide scanning along PSA. Exp Mol Pathol 2009; 86:36-40.
16. Lescar J, Stouracova R, Riottot MM, et al. Three-dimensional structure of an Fab-peptide complex: structural basis of HIV-1 protease inhibition by a monoclonal antibody. J Mol Biol 1997; 267:1207-22.
17. Bartonová V, Krá V, Sieglovál, et al. Potent inhibition of drug-resistant HIV protease variants by monoclonal antibodies. Antiviral Res 2008; 78:275-7.
18. Cavacini LA, Emes CL, Wisnewski AV, Power J, Lewis G, Montefiori D, Posner MR. Functional and molecular characterization of human monoclonal antibody. AIDS Res Hum Retroviruses 1998; 14: 1271-80.
19. Poumbourios P, McPhee DA, Kemp BE. Antibody epitopes sensitive to the state of human immunodeficiency virus type 1 gp41 oligomerization map to a putative alpha-helical region. AIDS Res Hum Retroviruses 1992; 8:2055-62.
20. Buchacher A, Predl R, Tauer C, et al. Human monoclonal antibodies against gp41 and gp120 as potential agents for passive immunization. Vaccines 1992; 92:191-5 .
21. Buchacher A, Predl R, Strutzenberger K, et al. Generation of human monoclonal antibodies against HIV-1 proteins, electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. AIDS Res Hum Retroviruses 1994; 10:359-69.
22. Gorny MK, Zolla-Pazner S. Human monoclonal antibodies that neutralize HIV-1. In HIV Immunology and HIV/SIV Vaccine Databases 2003. B.T.M. Korber, C. Brander, B.F. Haynes, R. Koup, J.P. Moore, B.D. Walker, D.I. Watkins, eds. Los Alamos National Laboratory, Theoretical Biology and Biophysics. Los Alamos, 2003; p. 37.
23. Nilsen BM, Haugan IR, Berg K, Olsen L, Brown PO, Helland DE. Monoclonal antibodies against human immunodeficiency virus type 1 integrase: epitope mapping and differential effects of integrase activities *in vitro*. J Virol 1996; 70:1580-7.
24. Sun NC, Ho DD, Sun CRY, et al. Generation and characterization of monoclonal antibodies to the putative CD4-binding domain of human immunodeficiency virus type 1 gp120. J Virol 1989; 63:3579-85.
25. Thornton JM, Edwards MS, Taylor WR, Barlow DJ. Location of 'continuous' antigenic determinants in the protruding regions of proteins. EMBO J 1986; 5: 409-13.
26. Sweet RM, Eisenberg D. Correlation of sequence hydrophobicities measures similarity in three-

dimensional protein structure. J Mol Biol 1983; 171: 479-88.
27. Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. Nature 1978; 275:673-4.
28. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions Proteins 1994; 19:141-9.
29. Koff WC, Kahn P, Gust I. Aids vaccine development. Challenges and opportunities. Caister Academic Press, 2007.
30. Kanduc D. Epitopic peptides with low similarity to the host proteome: towards biological therapies without side effects. Expert Opin Biol Ther 2009; 9: 45-53.