Identification and Characterization of Previously Described Epitopes in HIV-1 Subtypes B, C, F and BF in Brazil

Artur Trancoso Lopo de Queiroz[1,2], Luciane Amorim Santos[1,2], Domingos Ramon Moreau[1], Tulio de Oliveira[3], David I. Watkins[4], Bernardo Galvão-Castro[1,2] and Luiz Carlos Junior Alcantara[1,2]

[1]Advanced Public Health Laboratory, Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Salvador, BA, Brazil; [2]Bahia School of Medicine and Public Health/Foundation for Development of Science, Salvador, BA, Brazil; [3]Department of Zoology, University of Oxford, Oxford, United Kingdom; [4]University of Wisconsin Medical School, Department of Pathology, Madison, USA

Genetic analysis of HIV-1 is essential to improve treatment strategies and select epitopes for vaccine programs. The objective of this study was to determine whether known $CD4^+$ and $CD8^+$ epitopes were present in Brazilian HIV-1 strains. We used previously described $CD8^+$ and $CD4^+$ epitopes from the Los Alamos laboratory to search for these epitopes in the Brazilian sequences using the HIVbase program and we compared the frequency results with the analyses using physical-chemical profile tools from Network Protein Sequence Analysis (NPSA), and the SYFPEITHI program. Furthermore, this analysis was carried out with the Prosite tool using the GeneDoc program and ds/dn analyses using the Synonymous Nonsynonymous Analysis Program (SNAP). The HIVbase epitope mapping demonstrated that 30 $CD8^+$ and 6 $CD4^+$ epitopes were present in the Brazilian sequences at a high frequency. Only two of these epitopes were heavily glycosylated. Interestingly, ds/dn analyses showed evidence of purifying selective pressure. These types of analyses could be useful for the assessment of possible vaccine efficiency in populations.

Key-Words: HIV-1, *env*, vaccine, epitope.

HIV-1, identified as the etiological agent of AIDS [1], contributes to the development of immunodeficiency. Its biological complexity and high mutation rate have made the design of a vaccine to control the pandemic difficult. $CD8^+$ T lymphocyte responses are largely responsible for controlling viral replication during both acute and chronic infection [2-4]. The antibody responses appears much later and select a mutant virus [5]. Mutations in CTL epitopes and sites recognized by antibodies and acquisition of glycosylation sites are escape mechanisms that allow the virus to replicate and infect more cells [5-7].

The mapping of epitopes in diverse virus proteins and the identification of possible modifications can provide useful information to aid vaccine development. For example, mutations in the TW10 and SL9 Gag epitopes result in a fitness cost to the virus [7], and this kind of information is very useful for the vaccine development process. The use of bioinformatics software can identify escaping epitopes. Bioinformatics programs have been developed for diverse areas including protein analysis, physical-chemical characteristics, MHC binding databases, and posttranslational modification. The use of these programs to study the molecular epidemiology and genetic variation of the HIV-1 epidemic in Brazil may provide important information about the epidemic in this country. This information could be important in the design of an effective vaccine as well as for antiretroviral treatment. The objective of this study was to characterize genetic variation in previously identified epitopes in the Brazilian HIV-1 *env* sequences.

Received on 27 October 2006; revised 22 January 2007.

Address for correspondence: Dr. Luiz Carlos J. Alcantara. Advanced Public Health Laboratory, Gonçalo Moniz Research Center, Oswaldo Cruz Foundation. Rua Waldemar Falcão 121, Candeal, Salvador, Bahia, Brazil. Zip code: 40296-610. Telephone # 55 71 3176 2246/3176 2213. Fax # 55 71 3176 2300. E-mail: lalcan@cpqgm.fiocruz.br.

Financial Support: This project was partially supported by the Fundação de Amparo a Pesquisa do Estado da Bahia (FAPESB, grant number 303/03) and PN-DST/AIDS, Ministério da Saúde, Brasil (grant number 306 and 307/04).

The Brazilian Journal of Infectious Diseases 2007;11(1):27-30.

Materials and Methods

Sequences of all of the Brazilian HIV-1 strains (3,813) were collected from GenBank and added to the HIVbase Database [8]. Epitope mapping analyses were performed as described in the HIV Immunology and HIV/SIV Los Alamos Vaccine Databases [9]. We selected and analyzed only the *env* region, consisting of 2,644 sequences from gp120: C1V1=51; C2=59; C3=200; V3loop=515; and gp41=50. The alignment of these sequences was carried out using the ClustalX [10] software. Genedoc [11] software was used to edit and translate the alignment, and the potential site analyses were performed using the Prosite [12] tool. For selective pressure analysis, we used the Synonymous Nonsynonymous Analysis Program (SNAP) [13] from Los Alamos. The proportion of synonymous substitutions per potential synonymous site and the proportion of nonsynonymous substitutions per potential nonsynonymous site were calculated using the Nei and Gojobori method [14]. Prediction of cellular epitopes was made with the SYFPEITHI online database [15]. Identification of Antibody epitopes and their physical-chemical characteristics were carried out using the Physico-chemical profiles program of the Network Protein Sequence Analysis (NPSA) [16-20].

Results

HIVbase epitope mapping showed that thirty $CD8^+$ (Table 1) and six $CD4^+$ (Table 2) epitopes had a high frequency and showed varying degrees of conservation in the Brazilian HIV sequence. However, the SYPETEITHI analysis was restricted to the following HLA alleles: HLA-A*03, -A*6801, -A*2402 and -A*0201. Two of the epitopes in this program had several

Table 1. Mapping of the frequency of the Los Alamos CD8 epitopes in the Brazilian HIV-1 sequences

HXB2 Location Protein region	% Similarity	Epitopes sequence	Frequency	HLA	HXB2 Location Protein region	% Similarity	Epitopes sequence	Frequency	HLA
gp120/C1	Total=72.5	33NLWVTVYYGV42	37/51	A02	gp120 /C1	Total=98.0	34LWVTVYYGV42	50/51	A*0201
subtype B	30	. . . . .	3/10		subtype B	100	. . . . .	10/10	
		K.	4/10						
		Q.	3/10						
subtype F	100	. . . . .	10/10		subtype F	100	. . . . .	10/10	
subtype C	100	. . . . .	12/12		Subtype C	100	. . . . .	12/12	
B/F	63	. . . . .	12/19		B/F	94.7	. . . . .	18/19*	
recombinant		Q.	2/19		recombinant				
		K.	2/19						
		D.	2/19*						
gp 120/C1	Total=100	36VTVYYGVPV44	51/51	A02	gp 120/C1	Total=92.1	36VTVYYGVPVWK46	47/51	A*6801
Subtype B	100	. . . . .	10/10		Subtype B	100	. . . . .	10/10	
Subtype F	100	. . . . .	10/10		Subtype F	100	. . . . .	10/10	
Subtype C	100	. . . . .	12/12		Subtype C	83.3	. . . . .	10/12	
							. . . . . R	2/12	
B/F	100	. . . . .	19/19		B/F recombinant	89.5	. . . . .	17/19	
recombinant							. . . . . R	2/19	
gp 120/C1	Total=92.1	37TVYYGVPVWK46	47/51	A*0301, A*6801,	gp 120/C1	Total=90.2	38VYYGVPVWKEA48	46/51	Cw7
Subtype B	100	. . . . .	10/10		Subtype B	100	. . . . .	10/10	
Subtype F	100	. . . . .	10/10		Subtype F	90	. . . . .	9/10*	
Subtype C	83.3	. . . . .	10/12		Subtype C	83.3	. . . . .	10/12	
		. . . . . R	2/12				. . . . . R . .	2/12	
B/F	89.5	. . . . .	17/19		B/F recombinant	89.5	. . . . .	17/19	
recombinant		. . . . . R	2/19				. . . . . R . .	2/19	
gp 120/C1	Total=62.7	42VPVWKEATTT51	32/51	B*5501, B55	gp 120/C1	Total=62.7	42VPVWKEATTTL52	32/51	B*3501
Subtype B	80	. . . . .	8/10*		Subtype B	80	. . . . .	8/10*	
Subtype F	90	. . . . .	9/10*		Subtype F	90	. . . . .	9/10*	
Subtype C	0	. . . . . .K. .	10/12*		Subtype C	0	. . . . . .K . . .	10/12*	
B/F	78.9	. . . . .	15/19		B/F recombinant	78.9	. . . . .	15/19	
recombinant		. . . . R . . N. .	2/19*				. . . . R . . N . . .	2/19*	
gp 120/C1	Total=90.2	50TTLFCASDAK59	46/51	A3supertype	gp 120/C1	Total=92.1	51TLFCASDAK59	47/51	A3supertype
Subtype B	100	. . . . .	10/10		Subtype B	100	. . . . .	10/10	
Subtype F	100	. . . . .	10/10		Subtype F	100	. . . . .	10/10	
Subtype C	75	. . . . .	9/12		Subtype C	83.3	. . . . .	10/12	
		. . . . . .R	2/12*				. . . . . R	2/12	
B/F	89.5	. . . . .	17/19*		B/F recombinant	89.5	. . . . .	17/19*	
recombinant									
gp 120/C1	Total=90.2	108IISLWDQSL116	46/51	A2.1	gp 120/C1	Total=94.1	109ISLWDQSLK117	48/51	A11
Subtype B	100	.. . . . .	10/10		Subtype B	100	. . . . .	10/10	
Subtype F	100	.. . . . .	10/10		Subtype F	100	. . . . .	10/10	
Subtype C	83.3	.. . . . .	10/12*		Subtype C	91.6	. . . . .	11/12*	
B/F	84.2	.. . . . .	16/19		B/F recombinant	89.5	. . . . .	17/19*	
recombinant		V.	2/19*						
gp 120/C1V1	Total=94.1	110SLWDQSLKP118	48/51	A03	gp 120/C1V1	Total=92.1	117KPCVKLTPLC126	47/51	B7
Subtype B	100	. . . . .	10/10		Subtype B	100	. . . . .	10/10	
Subtype F	100	. . . . .	10/10		Subtype F	90	. . . . .	9/10*	
Subtype C	91.6	. . . . .	11/12*		Subtype C	100	. . . . .	12/12	
B/F	89.5	. . . . .	17/19*		B/F recombinant	84.2	. . . . .	16/19*	
recombinant									
gp120/C2	Total=61.0	252KPVVSTQLLL261	36/59	B07,B08	gp120/V3loop	Total=54.9	296CTRPNNNTRK305	283/515	A03;A02
Subtype B	28.6	. . . . .	4/14		Subtype B	47.7	. . . . .	187/392	
		R.	10/14				. . . .G	31/392	
Subtype F	100	. . . . .	11/11		Subtype F	84.4	. . . . .	38/45	
							. . . . . Y . .	2/45	
Subtype C	50	. . . . .	7/14		Subtype C	55.2	. . . . .	16/29	
		. . . I	3/14				. . . . . .E	8/29	
		. . . M.	2/14*				. . . .G	2/29	
B/F	70	. . . . .	14/20		B/F recombinant	85.7	. . . . .	42/49	
recombinant		R.	6/20				. . . S	2/49*	
gp120/C3	Total=60.5	375SFNCGGEFF383	121/200	B1516; B15;B63	gp120/C3	Total=60.0	375SFNCGGEFFY384	120/200	A29
Subtype B	68.4	. . . . .	117/153		Subtype B	68.4	. . . . .	116/153	
		T.	24/153				T.	23/153	
		. . . . A	4/153				. . . . A	4/153*	
		. . . . R . . .	2/153*						
Subtype F	0	. . . . . .R	10/14		Subtype F	0	. . . . . R .	10/14	
		N. . . M	4/14				N . . .M	4/14	
Subtype C	0	. . . . R	7/9*		Subtype C	0	. . . . R	7/9*	
B/F	16.6	. . . . .	4/24		B/F recombinant	16.6	. . . . .	4/24	
recombinant		. . . . R	18/24*				. . . . R	18/24*	
gp120/C3	Total=73.5	376FNCGGEFF383	147/200	Cw4	gp120/C3	Total=72.5	376FNCGGEFFY384	145/200	A29

Table 1. (continued)

Subtype B	89.5	. . . . .	141/153		Subtype B	88.6	. . . . .	139/153	
		. . . A	4/153				. . .A	4/153*	
		. . . R	2/153*						
Subtype F	0	. . . . . R	10/14		Subtype F	0	. . . . . R.	10/14	
		. . . M	4/14				. . .M	4/14	
Subtype C	0	. . . R	7/9*		Subtype C	0	. . .R	7/9*	
B/F	25	. . . . .	6/24		B/F recombinant	25	. . . . .	6/24	
recombinant		. . . R	18/24				. . .R	18/24	
gp120/C3	Total=67.5	377NCGGEFFYCN386	135/200	ND**	gp 41	Total=74.0	529TMGAASITL537	37/50	A2
Subtype B	88.6	. . . . .	129/153		Subtype B	28.6	. . . . .	2/7	
		. . . . . D	10/153				. . . . . L. .	2/7	
		. . A	4/153*				. . . . . VA.	2/7*	
Subtype F	0	. . . . . R . . .	9/14		Subtype F	90.9	. . . . .	10/11*	
		. . M	4/14*						
Subtype C	0	. . R	7/9*		Subtype C	100	. . . . .	13/13	
B/F	25	. . . . .	6/24		B/F recombinant	63.2	. . . . .	12/19	
recombinant		. . R	18/24				. . . . . M. .	3/19*	
gp 41	Total=52.0	565LLQLTVWGI573	26/50	A2	gp 41	Total=70.0	584ERYLKDQQL592	35/50	B14,A32
Subtype B	42.9	. . . . .	3/7		Subtype B	42.9	. . . . .	3/7	
		. . . . . .M	4/7				. . . .R . . .	2/7	
							. . . .G . . .	2/7	
Subtype F	100	. . . . .	11/11		Subtype F	63.6	. . . . .	7/11	
							. . . .Q . . .	4/11	
Subtype C	0	M.	13/13		Subtype C	73.9	. . . . .	10/13	
							. . . .R . . .	2/13	
B/F	63.2	. . . . .	12/19		B/F recombinant	78.9	. . . . .	15/19	
recombinant		M.	6/19*				. . . .Q . . .	3/19*	
gp 41	Total=70.0	584ERYLKDQQLLG594	35/50	ND**	gp 41	Total=70.0	585RYLKDQQLL593	35/50	A*23,A24
Subtype B	42.9	. . . . .	3/7		Subtype B	42.9	. . . . .	3/7	
		. . . . R	2/7				. . . R. . . .	2/7	
		. . . . G	2/7				. . . G. . . .	2/7	
Subtype F	63.6	. . . . .	7/11		Subtype F	63.6	. . . . .	7/11	
		. . . . Q	4/11				. . . Q. . . .	4/11	
Subtype C	73.9	. . . . .	10/13		Subtype C	73.9	. . . . .	10/13	
		. . . . R	2/13*				. . . R. . . .	2/13 *	
B/F	78.9	. . . . .	15/19		B/F recombinant	78.9	. . . . .	15/19	
recombinant		. . . . Q	3/19*				. . . Q. . . .	3/19*	
gp 41	Total=90.0	678WLWYIKIFI686	45/50	A2	gp 41	Total=82.0	680WYIKIFIMI688	41/50	A*2402
Subtype B	100	. . . .	7/7		Subtype B	100	. . . .	7/7	
Subtype F	90.9	. . . .	10/11*		Subtype F	90.9	. . . .	10/11*	
Subtype C	92.3	. . . .	12/13*		Subtype C	84.6	. . . .	11/13*	
B/F	84.2	. . . .	16/19		B/F recombinant	68.4	. . . .	13/19	
recombinant		. . . .R.. .	2/19*				. .R.. . . .	2/19	
							. . . .. I.	2/19*	
gp 41	Total=82.0	681YIKIFIMIV689	41/50	A2	gp 41	Total=60.0	846RIRQGLERA859	30/50	A*0205
Subtype B	100	. . ..	7/7		Subtype B	100	. . . . .	7/7	
Subtype F	90.9	. . ..	10/11*		Subtype F	63.6	. . . . .	7/11	
							. . . . F . . .	4/11	
Subtype C	84.6	. . ..	11/13*		Subtype C	0	. . . . F .A .	7/13	
							. . . . . A .	3/13*	
B/F	68.4	. . ..	13/19		B/F recombinant	84.2	. . . . .	16/19	
recombinant		. .R.. . . .	2/19				. . . . F. .	2/19*	
		. . .. . I . .	2/19*						

* Mutation with frequency lower than 5% was excluded from the table. ** HLA not determined. % similarity: % of sequences that have that epitope sequence. Frequency: nº of sequences/nº of sequences that have that epitope or mutation.

Table 2. Mapping of the frequency of the Los Alamos CD4 epitopes in the Brazilian HIV-1 sequences

HXB2 Location Protein region	% Similarity	Epitope sequency	Frequency	HLA	HXB2 Location Protein region	% Similarity	Epitope sequency	Frequency	HLA
gp 41	Total=50.0	562QQHLLQLTVWGIKQL576	25/50	ND**	gp 120 C1	Total=88.2	118IISLWDQSLKPC119	45/51	ND**
Subtype B	42.8	. . . . .	3/7		Subtype B	100	... .	10/10	
		. . . M.	4/7						
Subtype F	100	. . . . .	11/11		Subtype F	100	... .	10/10	
Subtype C	0	. . . M.	13/13		Subtype C	83.3	... .	10/12*	
B/F recombinant	57.9	. . . . .	11/19		B/F	78.9	... .	15/19	
		. . . M.	6/19*		recombinant		V.	2/19*	
gp 41	Total=54.0	593LGIWGCSGKLIC604	27/50	ND**	gp 120 C1	Total=92.1	110SLWDQSLKPCVKLTPL125	47/51	ND**
Subtype B	100	. . . . .	7/7		Subtype B	100	. . . . .	10/10	
Subtype F	0	. . L	11/11		Subtype F	100	. . . . .	10/10	
Subtype C	100	. . . . .	13/13		Subtype C	91.6	. . . . .	11/12*	
B/F recombinant	36.8	. . . . .	7/19		B/F	84.2	. . . . .	16/19*	
		. . L	9/19		recombinant				
		. . L . . . .R. .	2/19*						
gp 41	Total=54.0	594GIWGCSGKLIC604	27/50	ND**	gp 41	Total=54.0	594GIWGCSGKLI603	27/50	ND**
Subtype B	100	. . . . .	7/7		Subtype B	100	. . . . .	7/7	
Subtype F	0	. L	11/11		Subtype F	0	.L	11/11	
Subtype C	100	. . . . .	13/13		Subtype C	100	. . . . .	13/13	
B/F recombinant	36.8	. . . . .	7/19		B/F	36.8	. . . . .	7/19	
		. L	9/19		recombinant		.L	9/19	
		. L . . . . R. .	2/19*				.L . . . . R . .	2/19*	

*Mutation with frequency lower than 5% was excluded from the table. ** HLA not determined. % similarity: % of sequences that have the epitope. Frequency: nº of sequences/nº of sequences that have the epitope or mutation.

N-glycosylation sites (CTRPNNNTRK at amino acid position 296 to 305, at a frequency of 96.2%, and NCGGEFFYCN at amino acid position 377 to 386 with a frequency of 84.5%). The *ds/dn* ratio was high in some of the $CD8^+$ epitopes. This high ratio suggests that these particular epitopes may either not be under positive selection or are maintained under functional constraints.

The subtype B epitopes were the most conserved ones, especially in the C1 and C3 regions, followed by subtype F. In the V3loop region, subtype F was the most conserved one. However, none of the most frequent mutations were associated with the loss of N-glycosylation site at this position. The gp41 was the most conserved region among the subtypes and this can be explained by the absence of variable regions in this protein. Epitopes in this region showed high *ds/dn* ratios but these ratios were lower than the gp120 epitopes. This epitope conservation and the high *ds/dn* ratio suggest that these regions may be important for viral fitness. Mutations in this region might change the protein structure, reducing the infection capacity of the virus.

The epitope VPVWKEATTTL is associated with a rapid progression HLA allele, HLA-B35 [4] and exhibited low variation in non-C subtypes. Interestingly, this epitope was highly variable in B/F recombinants. This suggests that CTL may be exerting selective pressure on this subtype.

The regions involved in N-glycosylation were highly conserved. These sites are potentially important for the functioning of these proteins, and mutation in these regions might affect viral function.

Conclusion

An ideal vaccine would contain epitopes that would engender strong immune responses against the functionally important regions. Escape from these vaccine-induced immune responses would compromise viral fitness. Additionally, it will be important to continue analyzing epitope variability in other viral proteins in the HIV-1 strains circulating in Brazil as any eventual vaccine for use in Brazil will need to be relevant to the viruses in Brazil.

Acknowledgements

ATLQ and LAS contributed equally to this work. We are grateful to Elisabeth Deliege for her technical assistance.

References

1. Barré-Sinoussi F., Chermann J.C., Rey F., et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **1983**;220:868-71.
2. Matano T., Shibata R., Siemon C., et al. Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. J Virol **1998**;72:164-9.
3. Jin X., Bauer D.E., Tuttleton S.E., et al. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. J Exp Med **1999**;189:991-8.
4. Carrington M., Nelson G.W., Martin M.P., et al. HLA and HIV-1: heterozygote advantage and B35-Cw04 disadvantage. Science **1999**;283:1748-52.
5. Burton D.R., Stanfield R.L., Wilson I.A. Antibody *vs*. HIV in a clash of evolutionary titans. Proc Natl Acad Sci **2005**;102:14943-8.
6. Reitter J.N., Means R.E., Desrosiers R.C. A role for carbohydrates in immune evasion in AIDS. Nat Med **1998**;4:679-84.
7. Goulder P.J., Watkins D.I. HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol **2004**;4:630-40.
8. Lamers S., Beason S., Dunlap L., Compton R., Salemi M. HIVbase: a PC/Windows-based software offering storage and querying power for locally held HIV-1 genetic, experimental and clinical data. Bioinformatics **2004**;20:436-8.
9. Bette T. M. Korber, Christian Brander, Barton F. Haynes, et al. HIV Immunology and HIV/SIV Vaccine Databases. Los Alamos National Laboratory, Theoretical Biology and Biophysics. Los Alamos, New Mexico. **2003**.
10. Jeanmougin F., Thompson, J.D., Goy M., et al. Multiple sequence alignment with Clustal X. Trends Biochem Sci **1998**;23:403-5.
11. Nicholas K.B., Nicholas H.B.J., Deerfield D.W. GeneDoc: Analysis and Visualization of Genetic Variation. EMBNEW.NEWS **1997**;4:14.
12. Falquet L., Pagni M., Bucher P., et al. The PROSITE database. Nucleic Acids Res **2002**;30:235-8.
13. Korber B. HIV Signature and Sequence Variation Analysis. In: Allen G., Rodrigo, Learn G.H., eds. Computational Analysis of HIV Molecular Sequences. Dordrecht, Netherlands: Kluwer Academic Publishers, **2000**, p.55-72.
14. Nei M., Gojobori T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. Mol Biol Evol **1986**;5:418-26.
15. Rammensee H., Bachmann J., Emmerich N.P., et al. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics **1999**;50:213-9.
16. Hopp T.P., Woods K.R. A computer program for predicting protein antigenic determinants. Mol Immunol **1983**;20:483-9.
17. Kyte J., Doolittle R.F. A simple method for displaying the hydropathic character of a protein. J Mol Biol **1982**;157:105-32.
18. Karplus P.A., Schulz G.E. Prediction of chain flexibility in proteins. Naturwissens-chaften **1985**;72:212-3.
19. Parker J.M., Guo D., Hodges R.S. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry **1986**;25:5425-32.
20. Argos P., Rao J.K., Hargrave P.A. Structural prediction of membrane-bound proteins. Eur J Biochem **1982**;128:565-75.