Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design

Aline Cristina Mota-Miranda/*, Tulio de-Oliveira, Domingos Ramon Moreau, Catarina Bomfim***, Bernardo Galvão-Castro/*, Luiz Carlos Junior Alcantara/*/+**

Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Moniz-Fiocruz, Rua Valdemar Falcão 121, 40296-610 Salvador, BA, Brasil *Escola Bahiana de Medicina e Saúde Pública, Fundação Bahiana para o Desenvolvimento das Ciências, Salvador, BA, Brasil **MRC Pathogen Bioinformatics Unit, South African National Bioinformatics Institute, UWC, Cape Town, South Africa and Zoology Department, Oxford University, Oxford, United Kingdom ***Faculdade de Tecnologia e Ciência, Salvador, BA, Brasil

This study was carried out to evaluate the molecular pattern of all available Brazilian human T-cell lymphotropic virus type 1 Env (n = 15) and Pol (n = 43) nucleotide sequences via epitope prediction, physico-chemical analysis, and protein potential sites identification, giving support to the Brazilian AIDS vaccine program. In 12 previously described peptides of the Env sequences we found 12 epitopes, while in 4 peptides of the Pol sequences we found 4 epitopes. The total variation on the amino acid composition was 9 and 17% for human leukocyte antigen (HLA) class I and class II Env epitopes, respectively. After analyzing the Pol sequences, results revealed a total amino acid variation of 0.75% for HLA-I and HLA-II epitopes. In 5 of the 12 Env epitopes the physico-chemical analysis demonstrated that the mutations magnified the antigenicity profile. The potential protein domain analysis of Env sequences showed the loss of a CK-2 phosphorylation site caused by D197N mutation in one epitope, and a N-glycosylation site caused by S246Y and V247I mutations in another epitope. Besides, the analysis of selection pressure have found 8 positive selected sites (w = 9.59) using the codon-based substitution models and maximum-likelihood methods. These studies underscore the importance of this Env region for the virus fitness, for the host immune response and, therefore, for the development of vaccine candidates.

The human t-cell lymphotropic virus type 1 (HTLV-1) is the etiologic agent of adult T-cell leukemia lymphoma (ATLL) (Poiesz et al. 1980) and tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM) (Gessain et al. 1985, Osame et al. 1986). This infection is endemic in Japan, in the Caribbean Basin, some South American, and African regions (Mueller 1991), while there are 2.5 million HTLV-1 infected people in Brazil (Carneiro-Proietti et al. 2002). In contrast, only 5% of infected people develop the associated disease, and it is not yet known why 95-98% of them remain asymptomatic. The envelope glycoproteins of this virus are known to play a critical role in the infection process. The surface glycoprotein (gp46) subunit is involved in cellular receptor recognition, while the transmembrane glycoprotein (gp21) subunit anchors the gp46 to the cell and plays a major role in the post-binding steps of the fusion process, resulting in the formation of multinucleated giant cells named syncytia. These envelope proteins are expressed on the surface of virus-infected cells and on viral particles and they are the first to be recognized by the host in the course of the natural immune response (Nagy et al. 1983, Palker et al. 1989), in such a way that HTLV-1 infected individuals develop a strong immune response to the envelope (*env*) gene products. Several studies have focused on mapping linear immunodominant regions of the HTLV-1 surface glycoprotein that could elicit antibody responses, for the purpose of vaccine development and diagnostic screening (Tanaka et al. 1991, Inoue et al. 1992, Kuroki et al. 1992, Baba et al. 1993, Desgranges et al. 1994). Specific antibodies against envelope determinants can inhibit HTLV-1-mediated syncytium formation, suggesting possible protection from the infection cycle (Hadlock et al. 2002). Multiple neutralizing monoclonal antibodies to linear epitopes within amino acids 175-200 of the gp46 and more rarely to epitopes into the carboxy terminal of the protein have been described using immunoreactive assays (Baba et al. 1993). The antiviral cytotoxic T-lymphocyte (CTL) is activated in vivo and directed mainly to the HTLV-1 Tax regulatory protein and to a lesser extent to the structural Gag, Env, and Pol proteins (Kannagi et al. 1991, Parker et al. 1992, Elovaara et al. 1993, Pique et al. 1996).

Possible strategies for vaccine development should evaluate the antigenic capacity of viral peptides in stimulating the humoral and cellular immune responses. Therefore, a vaccine development process would be more specific if the practice of screening important genomic re-

gions would be carried out. This study was therefore developed to evaluate the molecular pattern of all available Brazilian HTLV-1 Env and Pol sequences concerning to: physico-chemical analysis, to infer the peptide capacity of generating functional antibodies; protein potential site analysis, to get information of a possible immune system escape mechanism, such as changes in the protein structure; and epitopes prediction, to identify the possible targets for B and T-cells through human leukocyte antigen (HLA) molecules presentation. Furthermore, it is important to make sure that bioinformatic tools are safe and precise for screening new peptides to be tested in vitro.

MATERIALS AND METHODS

Sequences selection - All sequences analyzed in this study were selected from the GenBank, between March and May 2006. There were 44 available *pol* gene nucleotide sequences in the GenBank, but one of them (U12108) were rejected from the analysis because it was a too short (140 pb) fragment. At that time also, there were about 59 available *env* gene nucleotide sequences, but 44 of them were rejected from the analysis because they were isolated from Argentine, Caribbean, Romanian, and South African infected individuals, and because they were fragments with less than 500 pb. So, we have analyzed 15 *env* gene sequences, corresponding to the nucleotide 535 of gp46 to 153 of the gp21, and 43 *pol* sequences, from nucleotide 2,058 to 2,375, that represent all *env* and *pol* virus strains isolated from Brazilian individuals until May, 2006. To perform all analyses, two alignments (*env* and *pol*) were carried out with the ATK1 reference strain (Seiki et al. 1983) using the Clustal X 1.83 software (Jeanmougin et al. 1998).

All *env* nucleotide sequences previously deposited in the GenBank and used in the study are listed below with their corresponding accession number: PT9ATL (U81869); PT12ATL (U81865); PT3ATL (U81866); PT5ATL (U81867); PT8ATL (U81868). The new nucleotide sequences are: *env* - IDUSSA (DQ007198); FNN159 (DQ007197); FNN155 (DQ007194); FNN148 (DQ007191); FNN149 (DQ007197); FNN153 (DQ007193); FNN158 (DQ007158); FNN09 (DQ007189); FNN100 (DQ007190); FNN156 (DQ007195); *pol* - G0110 (AF197327); TP98/70 (AF197326); TP98/66 (AF197325); TP98/55 (AF197324); TP97/38 (AF197323); TP97/35 (AF197322); TP97/33 (AF197321); RNTP97/38 (AF197320); FNN094 (AF197319); FNN091 (AF197318); FNN089 (AF197317); FNN087 (AF197316); FNN083 (AF197315); FNN082 (AF197314); FNN081 (AF197313); FNN080 (AF197312); FNN078 (AF197311); FNN073 (AF197310); FNN072 (AF197309); FNN071 (AF197308); FNN068 (AF197307); FNN064 (AF197306); FNN061 (AF197305); FNN060 (AF197304); FNN057 (AF197303); FNN054 (AF197302); FNN053 (AF197301); FNN051 (AF197300); FNN047 (AF197299); FNN043 (AF197298); FNN041 (AF197297); FNN039 (AF197296); FNN035 (AF197295); FNN032 (AF197294); FNN029 (AF197293); FNN028 (AF197292); FNN026 (AF197291); FNN022 (AF197290); FNN019 (AF197289); FNN009 (AF197288); G0263 (AF197287); G0153 (AF197286); TP98/75 (AF197285).

Epitopes prediction - Before performing the epitope prediction, we have screened from the literature, 12 and 4 peptides previously described and tested in vitro (Horal et al. 1991, Baba et al. 1993, Pique et al. 1996, Schonbach et al. 1996), corresponding to the *env* and *pol* gene, respectively. The main criterion used to screen the peptides were their frequency in the papers. The epitope prediction was carried out for all Env and Pol amino acid sequences to 14 HLA-I (HLA A26, HLA A1510, HLA A4402, HLA A01, HLA A0201, HLA A2402, HLA A5101, HLA A03, HLA A0203, HLA A2705, HLA A6801, HLA A08, HLA A0702, HLA A2709) and 6 HLA-II alleles (HLA DRB1 0101, HLA DRB1 0401, HLA DRB1 0301, HLA DRB1 1501, HLA DRB1 0701, HLA DRB1 1101), using the online bioinformatics tool SYFPEITHI (Rammensee et al. 1999) (http://www.syfpeithi.de/Scripts/MHCServer.dll/Epitope Prediction.htm). This tool uses an algorithm that can predict sequences that have the potential ability to bind to one or more different HLA-I and HLA-II molecules. It also provides information about the epitope sequence, the specificity to the HLA molecule and the HLA binding score for each epitope. We performed the epitope prediction to the HLA alleles available in the software which are the most frequent in the database. After epitope prediction, similarity and variation calculations were performed comparing the peptide and epitope amino acid sequences within the Brazilian sequences, to get information about the specificity of the epitope binding to the HLA molecule and a possible epitope consensus sequence among different isolates.

Physico-chemical and potential protein domain analysis - To investigate possible influences of mutations, described into the *env* gene epitopes, we performed the physico-chemical analysis of the Env sequences using the Network Protein Sequence Analysis (NPSA) (Argos et al. 1982, Kyte & Doolittle 1982, Hopp & Woods 1983, Karplus & Schulz 1985, Parker et al. 1986) (http://npsa-pbil.ibcp.fr/) and the potential protein domain analysis using the GeneDoc software (Nicholas et al. 1997) and the Prosite tool, as previously described (Queiroz et al. 2007).

Selective pressure test - To test the hypothesis that the amino acid substitutions within the predicted epitopes into the *env* gene could have been favored or not by natural selection, the positive selection was assessed using six different codon-based maximum-likelihood (ML) substitution models (Yang et al. 2000). All models were implemented in the Codeml program of the PAML software package (Yang 1997). In agreement with a previously described study (de Oliveira et al. 2004), the ω and *p* values were estimated through maximum-likelihood optimization, in such a way that using the M3 model, sites with a posterior probability exceeding 90% and a

ω value > 1 were designated as being "positive selection sites". Finally, likelihood ratio test (LRT) analysis was used to determine: (1) if site heterogeneity selection was present and (2) if there were to be positive selection sites.

RESULTS

Epitopes prediction - Based on the peptide screening study, we identified the presence of 12 previously described peptides in all Brazilian Env sequences. These peptides cover the entire studied gp46 fragment and their respective locations in the protein according to ATK1 reference sequence are listed below (Table I). The total variation between the published peptide amino acid sequence and its composition in the Brazilian isolates was 17%, confirming that this viral protein region has too conservative characteristics. To search for possible epitopes into the peptides, we used SYFPEITHI tool. From our results, it was possible to identify one epitope in each studied peptide, demonstrating that the screened peptides are true targets for host immune response, through the recognition of an HLA-I or HLA-II specific binder. To investigate the epitopes specificity, among viral isolates, we have calculated the similarity of the amino acid epitope sequence in all Brazilian studied strains. Our results indicated that the observed mutations resulted in a total variation of 9 and 17% for HLA-I and HLA-II specific epitopes, respectively. The difference between the total variation of HLA-I and HLA-II epitopes can be explained by the fact that the typical length of HLA-I ligand comprises 9aa, while the length of an HLA-II ligand is much longer (15aa). Interestingly, both Env and Pol predicted epitopes, through the computer analysis, were able to bind more than one HLA allele, demonstrating a promiscuous behavior of those antigenic determinants and also suggesting an approach of these results to the biological significance.

A similar set of analyses such as peptide screening, epitopes prediction, and similarity/variation calculations were also carried out for the 43 Pol amino acid sequences. Only the 4 most frequent peptides previously described (Pique et al. 1996, Schonbach et al. 1996) were screened to the Pol fragment analysis because the analyzed Brazilian sequences are just 172 amino acids longer. In such a way, all four published peptides were found in all Brazilian isolates and revealed conservative characteristics, through the amino acid peptide variation of 0.75%, which was much lower than the *env* gene percentages (Table II). When we performed the epitope prediction on the Pol peptides sequences, we found 4 HLA-I specific epitopes and 4 HLA-II specific epitopes, both of them with a very low rate of total epitope variation (0.75%) (Table II).

Physico-chemical and potential protein domain analysis - The physico-chemical analysis demonstrated that the entire studied fragment (gp46-gp21) is characterized by high antigenicity, flexibility, and hidrophilicity profiles. Interestingly, the results demonstrated that the mutations present in 5 of 12 epitopes had increased the antigenicity profile. When subjecting these epitopes to potential protein domain analysis to investigate if the mutations in the epitope had changed any signature into the protein (Table I), we observed that the D197N mutation in the first epitope (195-207aa) was the cause for the loss of a CK2- phosphorylation site (Fig. 1A). Similarly, the tenth epitope (238-250aa) which was charac-
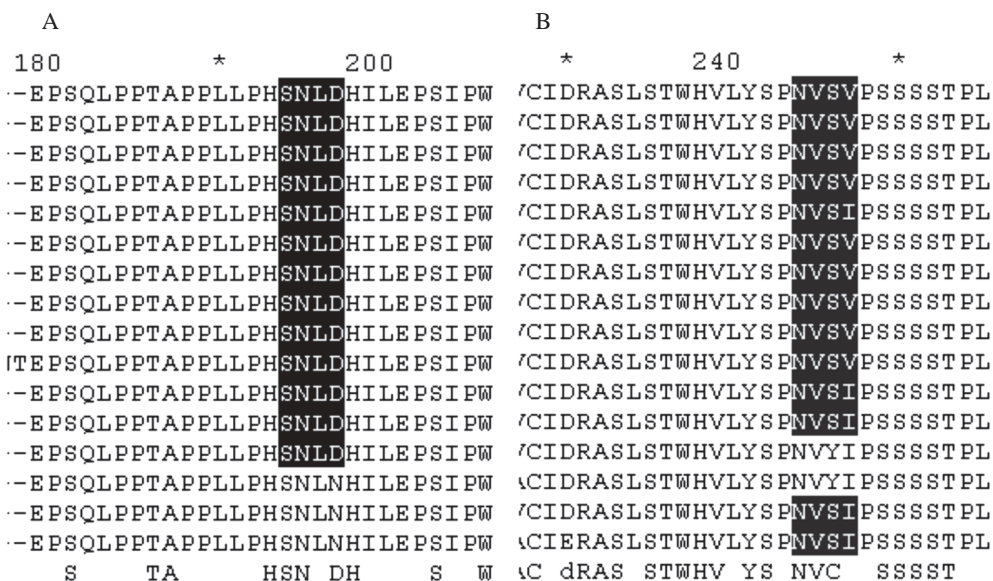

Fig. 1: potential protein domain analysis: loss of CK2-phosphorylation (A) and N-glycosylation (B) sites because of the D197N and (S246Y and V247I) mutations, respectively.

TABLE I

Human T-cell lymphotropic virus type 1 previous published Env (gp46) peptides versus possible human leukocyte antigen (HLA) class I and II epitopes and similarity with Brazilian sequences

HLA class	Peptide sequence	Similarity %		SYFPEITHI epitope prediction		
				Peptide sequence	Similarity %	HLA allele
	197-D H I L E P S I P W K S K L L L T L V Q L-216 N 3/15ᵃ	80		195-N L D H I L E P S I P W K-207 N 3/15ᵇ	80	1/2/3
	213-L V Q L T L Q S T N Y T C I V C I D R A S L S T-236 .G . . . . 1/15ᵃ .G I . . . . A . . . . 1/15ᵃ . . . . A 1/15ᵃ	80		211-L T L V Q L T L Q S T N Y-223 . . G . . . . 2/15ᵇ . . . . I . 1/15ᵇ	80	3/4/5/6/7/8
	235-S T W H V L Y S P N V S V P S S S S T P-254 . . . . Y I . . . . . . 2/15ᵃ . . . . I 5/15ᵃ	53		233-S L S T W H V L Y S P N V-245 . . . . 15/15ᵇ	100	◄
	253-T P L L Y P S L A L P A P H L T L P F N W T H C F D P Q I Q -282 . . . . 15/15ᵃ	100		251-S S T P L L Y P S L A L P-263 . . . . 15/15ᵇ	100	◄
	277-F D P Q I Q A I V S S P C H N S L-293 . . . . S 1/15ᵃ	93		275-C F D P Q I Q A I V S S P-287 . . . . 15/15ᵇ	100	◄
	288-P C H N S L I L P P F S L S P V P L G S R S R R A V P V A-317 S 15/15ᵃ	93		286-S S P C H N S L I L P P F-298 . S 1/15ᵇ	93	2/3/7/8/9/10
	176-L N T E P S Q L P P T A P P L L P H S N L D H I -199 – – . . . . N . . . 3/15ᵃ	80		179-E P S Q L P P T A P P L -190 . . . . 15/15ᵇ	100	11/12/13
	190-L L P H S N L D H I L E P S I P W K S K -209 . . . . N 3/15ᵃ	80		188-P P L L P H S N L D H I L-200 . . . . N . . . 3/15ᵇ	80	1/2/3/4/8/9/10
I	224-Y T C I V C I D R A S L S T W H V L Y S P -244 . . . . A . . . . 1/15ᵃ . . . . A . E 1/15ᵃ	86		222-N Y T C I V C I D R A S L-234 . . . . A . . . 1/15ᵇ . . . . A . E . 1/15ᵇ	85	11/12/13
	240-V L Y S P N V S V P S S S S T P L L Y P S L A-262 . . . . Y I . . . . 2/15ᵃ . . . . I 5/15ᵃ	53		238-H V L Y S P N V S V P S S-250 . . . . Y I . . 1/15ᵇ . . . . I . . 1/15ᵇ	85	1/2/3/4/6/8/9/10
	274-W T H C F D P Q I Q A I V S S P C H N S L I L-296 . . . . S 1/15ᵃ	93		272-N W T H C F D P Q I Q A I-284 . . . . 15/15ᵇ	100	2/3/4/7/9/10/11 12/13
	292-N S L I L P P F S L S P V P T L G S R S R R A-314 . . . . 15/15ᵃ	100		290-H N S L I L P P F S L S P-302 . . . . 15/15ᵇ	100	◄

Total peptide variation into the sequences: 17%   Total epitope variation into the sequences: 9%

197-D H I L E P S I P W K S K L L T L V Q L-216
    N   . 3/15 *a*

195-N L D H I L E P S I P W K S K L L T L-213
    N   . . 3/15 *b*   80   1/2/3/4

213-L V Q L T L Q S T N Y T C I V C I D R A S L S T-236
    G   . 1/15 *a*
    G . I   . 1/15 *a*
          . . . A . E   . 1/15 *a*
          . . . A   . 1/15 *a*

211-L T L V Q L T L Q S T N Y T C I V C I-229
    G   . 1/15 *b*
    G . I   . 1/15 *b*
          . . . A   . 1/15 *b*   80   1/4/5/3

235-S L S T W H V L Y S P N V S V P S S S S T P-254
                          Y I   . 2/15 *a*
                          . I   . 5/15 *a*

233-S L S T W H V L Y S P N V S V P S S S-251
                         Y I   . 5/15 *b*
                         . Y I   . 2/15 *b*   53   4/5/3

253-T P L L Y P S L A L P A P H L T L P F N W T H C F D P Q I Q-282
                                                          N . 15/15 *a*

251-S T P L L Y P S L A L P A P H L T L -269
                                   . 15/15 *b*   100   ▲▲

277-F D P Q I Q A I V S S P C H N S L-293
                    S   . 1/15 *a*

275-H C F D P Q I Q A I V S S P C H N S L-293
                     S   . 1/15 *b*   93   1/2/5/6

288-P C H N S L I L P F F S L S P V P L G S R S R R A V P V A-317
    S   . 15/15 *a*

286-S S P C H N S L I L P F F S L S P V P-304
      S   . 1/15 *b*   93   1/2/4/5/6/3

176-L N T E P S Q L P P T A P P L L P H S N L D H I -199
    – – –   N . 3/15 *a*

179-E P S Q L P P T A P P L L P H S N L-196
                                   . 15/15 *b*   100   ▲▲

190-L L P H S N L D H I L E P S I P W K S K -209
                     N   . 3/15 *a*

188-P L L P H S N L D H I L E P S I P W-206
                       N   . 3/15 *b*   80   ▲▲

224-Y T C I V C I D R A S L S T W H V L Y S P -244
              . A   . 1/15 *a*
              . A . E   . 1/15 *a*

222-N Y T C I V C I D R A S L S T W H V L 240
               . A   . 1/15 *b*
               . A . E   . 1/15 *b*   86   1/24/5/6/3

240-V L Y S P N V S V P S S S S T P L L Y P S L A-262
                 Y I   . 2/15 *a*
                 . I   . 5/15 *a*

238-H V L Y S P N V S V P S S S S T P L L-256
                   Y I   . 2/15 *b*
                   . I   . 5/15 *b*   53   1/2/3/4/5/6

274-W T H C F D P Q I Q A I V S S P C H N S L I L-296
                          S   . 1/15 *a*

272-N W T H C F D P Q I Q A I V S S P C H-290
                           S   . 1/15 *b*   93   1/2/3/4/6

292-N S L I L P F F S L S P V P T L G S R S R R A-314
                                     . 15/15 *a*

290-H N S L I L P F F S L S P V P T L G S-308
                                   . 15/15 *b*   100   ▲▲

Total peptide variation into the sequences: 17%   Total epitope variation into the sequences: 17%

TABLE II

Human T-cell lymphotropic virus type I previous published Pol peptides versus possible human leukocyte antigen class I and II epitopes and similarity with Brazilian sequences

MHC	Peptide sequence	Similarity %	SYFPEITHI epitope prediction		
			Epitope sequence	Similarity %	HLA allele
I	704- S L L Q A I A Y L -712 . . . . 43/43[a]	100	702-I S S L L Q A I A Y L G K -714 . . . . 43/43[b]	100	▲
	777-L P M D N A L S I -785 . S . . . . 1/43[a] . . I . . . . . . 1/43[a]	97	775-P D L P M D N A L S I A L -787 . . . S . . . . 1/43[b] . . . . I . . . . . . 1/43[b]	97	2/3/4/6/7/8/10 11/12/13
	709-I A H L G K P S Y -717 . . . . 43/43[a]	100	707-Q A I A H L G K P S Y I N -719 . . . . 43/43[b]	100	▲
	738-L A I R H T T H V P Y -748 . . . . 43/43[a]	100	736-T S L A I R H T T H V P Y N P -750 . . . . 43/43[b]	100	▲
	Total peptide variation into the sequences: 0.75%		Total epitope variation into the sequences: 0.75%		
II	704- S L L Q A I A Y L -712 . . . . 43/43[a]	100	702-I S S L L Q A I A Y L G K P S Y I N T -720 . . . . 43/43[b]	100	▲▲
	777-L P M D N A L S I -785 . S . . . . 1/43[a] . . I . . . . . . 1/43[a]	97	775-P D L P M D N A L S I A L W T I N -791 . . . S . . . . 1/43[b] . . . . I . . . . 1/43[b]	97	2/3/4/5/6/7/8 10/11/12/13
	709-I A H L G K P S Y -717 . . . . 43/43[a]	100	707-Q A I A H L G K P S Y I N T D N G P A -725 . . . . 43/43[b]	100	▲▲
	738-L A I R H T T H V P Y -748 . . . . 43/43[a]	100	736-T S L A I R H T T H V P Y N P T S S G L V -750 . . . . 43/43[b]	100	▲▲
	Total peptide variation into the sequences: 0.75%		Total epitope variation into the sequences: 0.75%		

a: peptide mutation frequency in all Brazilian sequences (n=43); *b*: epitope mutation frequency in all Brazilian sequences (n = 43); HLA I alleles: 1(HLA A 1510), 2 (HLA A 5101), 3 (HLA A 0201), 4 (HLA A 03), 5 (HLA A 01), 6 (HLA A 6801), 7 (HLA A 026), 8 (HLA 2705), 9 (HLA A 2402), 10 (HLA A 2709), 11 (HLA A 4402), 12 (HLA A 0702), 13 (HLA A 08) and 14 (HLA A 0203); HLA II alleles: 1 (HLA DRB1 0101), 2 (HLA DRB1 1101), 3 (HLA DRB1 0701), 4 (HLA DRB1 0301), 5 (HLA DRB1 1501) and 6 (HLA DRB1 0401); ▲ represent all HLA-I and ▲▲ HLA-II alleles.

terized by both S246Y and V247I mutations did not show the N-glycosylation site present in the epitope without these mutations (Fig. 1B).

Selective pressure test - ML methods were used to asses amino acid variation and identify targets of positive selection. The faction of sites (p_1, p_2, p_3), with the respective w (*dn/ds*) values for each model (M0-M3, M7-M8) are presented in the Table III. Although the env gene is considered to be too conservative, it was possible to identify 8 sites (codons) which are positive selected (w = 9.59) using models M2-M3 and M8. The LRT comparing the models support these results (p < 0,005).

DISCUSSION

In the present study, we have identified a set of multiple HLA-A and HLA-DR binding CD8 and CD4 T-cell epitopes, respectively, derived from the computer analysis performed at the SYFPHEITHI tool. The identified epitopes are conserved motifs into the envelope protein of viral isolates, since the variation of amino acid in the epitopes sequences was less than 20%. The knowledgement about the epitope sequence variation has great importance because the similarity between the predicted epitopes could be very useful to design an efficient vaccine for most common viral isolates. Furthermore, other studies using sera from HTLV-1-infected individuals have identified mutations within amino acids 175-200 that can affect antibody-mediated neutralization of HTLV-1 gp46-mediated syncytium formation (Blanchard et al. 1999, Tallet et al. 2001). Moreover, the observation of which alleles could bind to the mutant epitope is important, because the epitope containing the wild anchor motif could bind better to HLA molecules than other mutant motifs. Identifying which HLA allele the epitope is specific for could be also very useful to direct vaccine design for a specific population whose most frequent alleles have the strongest epitope's binder. Moreover, the computer epitope prediction is already used to the engineering of an HIV vaccine, confirming the utility of bioinformatics tools to select and construct novel "im-

TABLE III

Human T-cell lymphotropic virus type 1 Env selective pressure test: parameter estimated under six models of variable w (d_n/d_s) among sites

env gene	M0	M1	M2	M3	M7	M8	LRT I	LRT II
	$lnL = -864.96$ $w = 0.72$	$nL = -863.14$ $lp_1 = 0.61$ $p_2 = 0.39$ $w_1 = 0.0$ $w_2 = 1.0$	$lnL = -859.36$ $p_1 = 0.91$ $p_2 = 0.0$ $p_3 = 0.08$ $w_1 = 0.0$ $w_2 = 1.0$ $w_3 = 9.59$	$lnL = -859.36$ $p_1 = 0.91$ $p_2 = 0.0$ $p_3 = 0.08$ $w_1 = 0.0$ $w_2 = 1.0$ $w_3 = 9.59$	$lnL = -863.81$ $p_1 = 0.33$ $p_2 = 0.33$ $p_3 = 0.33$ $w_1 = 0.0$ $w_2 = 1.0$ $w_3 = 1.0$	$lnL = -859.36$ $p_1 = 0.30$ $p_2 = 0.30$ $p_3 = 0.30$ $p_4 = 0.08$ $w_1 = 0.0$ $w_2 = 0.0$ $w_3 = 0.0$ $w_4 = 9.59$	11.2 (p < 0.05)	7.56 (p < 0.05)
Selected sites (P = 1.0)			197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P	197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P		197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P		

The location of the selected codons is described according to HTLV-1 ATK1 reference sequence. LTR results support the presence of positive selection sites (p < 0.05) in the *env* gene.

munogenic consensus sequence" T cell epitopes for a globally relevant vaccine against HIV (De Groot et al. 2005, Fonseca et al. 2006).

The epitope prediction performed in this study is based on comparisons of precursor peptide sequences known to contain epitopes. It is the discovery of allele-specific motifs shared by eluted natural MHC ligands that allow the exact prediction of peptides from a given protein sequence. Every HLA allele has its individual peptide specificity that is defined by the position and specificity of the anchoring and non-anchoring amino acid residues. SYFPEITHI uses motifs matrices deducted from refined motifs based on a given pool of sequences and a single peptide analysis to search for natural ligands.

Although, these data had been obtained in computer analysis, our results are in agreement with experimental studies that have selected Tax and Env peptides to induce secretion of gamma interferon in peripheral blood mononuclear cells obtained from monkeys chronically infected with HTLV-1. After immunization, a high titre of antibodies and a high frequency of IFN-δ-producing cells were detected against the Env and the Tri-Tax immunogens, but not to the individual Tax peptides (Kazanji et al. 2006).

In line with previous reports, the post-transcriptional modifications, especially the glycosylation events are related to the formation of neutralizing antibodies, what explain our intention to compare the epitopes prediction results with carbohydrate moieties. Besides, there is an effort to engineer conformationally dependent HTLV-1 epitopes that elicit a protective immune response, especially a specific immune recognition and antibody responses (Kaumaya et al. 1995).

The physico-chemical analysis demonstrated that the entire studied fragment (gp46-gp21) had high antigenicity, flexibility and hidrophilicity profiles, and that mutations in that region could be important to the conformational epitope. This strategy is already investigated by previous studies, which report that peptide vaccines, to be able to induce high affinity and protective neutralizing antibodies, must rely in part on the design of antigenic epitopes, that mimic the three-dimensional structure of the corresponding region in the native protein (Sundaram et al. 2004).

It is known that a single-nucleotide mutation into an immunodominant CTL epitope can lead to viral escape and increased viral replication in HIV studies. Although there are no studies referring this experience with HTLV-1 isolates, two sites under positive selective pressure in our analysis were also responsible for the loss of CK2-phosphorylation and N-glycosylation moieties. This finding is very important to underscore the importance of the *env* gene to the immune response, and to confirm that this region is hardly submitted to the selective pressure, and also to the viral fitness.

Concerning that the proteins encoded by *pol* gene are not strong immune response inducers and that they are neither under positive selection nor an antibody response target, the Pol peptides were not submitted to physicochemical analysis nor potential domains.

This study was conducted with all available viral sequences isolated from Brazilian infected individuals with the objective to obtain a national panel of this kind of molecular investigation. In fact, there are not many Brazilian sequences of these regions available in the Gen Bank, but at least all of information here showed about epitope prediction, physico-chemical and protein potential site analysis, are new and could give support to future studies. We have not analyzed tax protein, because

there were only five available sequences in the GenBank at May 2006, and they corresponded to a short genomic region, but next studies should avaluate the tax protein, because the HTLV-1 tax protein is therefore a source of naturally processed epitopes presented by both HLA-I and HLA-II molecules, and are capable of stimulating both CD4+ and CD8+ T cell responses directed against tumors in vivo (Hanabuchi et al. 2001).

Nowadays, bioinformatics tools are being used to screen some immunodominant regions before testing in vitro, to get information about the genetic dynamics of the virus isolates and to monitor epidemic changes, such as their implications into the treatment response. Therefore, computer-aided analysis of protein sequences using various correlates of protein antigenicity is an important tool to the study of antigenic determinants.

REFERENCES

Argos P, Rao JK, Hargrave PA 1982. Structural prediction of membrane-bound proteins. *Eur J Biochem 128*: 565-575.

Baba E, Nakamura M, Tanaka Y, Kuroki M, Itoyama Y, Nakano S, Niho Y 1993. Multiple neutralizing B-cell epitopes of human T-cell leukemia virus type 1 (HTLV-1) identified by human monoclonal antibodies. A basis for the design of an HTLV-1 peptide vaccine. *J Immunol 151*: 1013-1024.

Blanchard S, Astier-Gin T, Tallet B, Moynet D, Londos-Gagliardi D, Guillemain B 1999. Amino acid changes at positions 173 and 187 in the human T-cell leukemia virus type 1 surface glycoprotein induce specific neutralizing antibodies. *J Virol 73*: 9369-9376.

Carneiro-Proietti AB, Ribas JG, Catalan-Soares BC, Martins ML, Brito-Melo GE, Martins Filho OA, Pinheiro SR, Araújo AQ, Galvão-Castro B, de Oliveira MS, Guedes AC, Proietti FA 2002. Infection and disease caused by the human T cell lymphotropic viruses type I and II in Brazil. *Rev Soc Bras Med Trop 35*: 499-508

De Groot AS, Marcon L, Bishop EA, Rivera D, Kutzler M, Weiner DB, Martin W 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine 23*: 2136-2148.

De Oliveira T, Salemi M, Gordon M, Vandamme AM, Van Rensburg EJ, Engelbrecht S, Coovadia HM, Cassol S 2004. Mapping sites of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design? *Genetics 167*: 1047-1058.

Desgranges C, Souche S, Vernant JC, Smadja D, Vahlne A, Horal P 1994. Identification of novel neutralization-inducing regions of the human T cell lymphotropic virus type I envelope glycoproteins with human HTLV-I-seropositive sera. *AIDS Res Hum Retroviruses 10*: 163-173.

Elovaara I, Koenig S, Brewah AY, Woods RM, Lehky T, Jacobson S 1993. High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. *J Exp Med 177*: 1567-1573.

Fonseca SG, Coutinho-Silva A, Fonseca LAM, Segurado AC, Moraes SL, Rodrigues H, Hammer J, Kallás EG, Sidney J, Sette A, Kalil J, Cunha-Neto E 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by the HIV-1 infected patients. *AIDS 20*: 2263-2273.

Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de The G 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet 2*: 407-410.

Hadlock KG, Yang Q, Rowe J, Foung SK 2002. Epitope mapping of human monoclonal antibodies recognizing conformational epitopes within HTLV type 1 gp46, employing HTLV type 1/2 envelope chimeras. *AIDS Res Hum Retroviruses 18*: 57-70.

Hanabuchi S, Ohashi T, Koya Y, Kato H, Hasegawa A, Takemura F, Masuda T, Kannagi M 2001. Regression of human T-cell leukemia virus type I (HTLV-I)-associated lymphomas in a rat model: peptide-induced T-cell immunity. *J Natl Cancer Inst 93*: 1775-1783.

Hopp TP, Woods KR 1983. A computer program for predicting protein antigenic determinants. *Mol Immunol 20*: 483-489.

Horal P, Hall WW, Svennerholm B, Lycke J, Jeansson S, Rymo L, Kaplan MH, Vahlne A 1991. Identification of type-specific linear epitopes in the glycoproteins gp46 and gp21 of human T-cell leukemia viruses type I and type II using synthetic peptides. *Proc Natl Acad Sci 88*: 5754-5758.

Inoue Y, Kuroda N, Shiraki H, Sato H, Maeda Y 1992. Neutralizing activity of human antibodies against the structural protein of human T-cell lymphotropic virus type I. *Int J Cancer 52*: 877-880.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ 1998. Multiple sequence alignment with clustal X. *Trends Biochem Sci 23*: 403-405.

Kannagi M, Harada S, Maruyama I, Inoko H, Igarashi H, Kuwashima G, Sato S, Morita M, Kidokoro M, Sugimoto M 1991. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8$^+$ cytotoxic T cells directed against HTLV-I-infected cells. *Int Immunol 8*: 761-767.

Karplus PA, Schulz GE 1985. Prediction of chain flexibility in proteins. *Naturwissens-chaften 72*: 212-213.

Kaumaya PT, Conrad SF, DiGeorge AM, Lairmore MD 1995 Glycosylation-dependent peptide antigenic determinants of env gp46 HTLV-1. *Leukemia 9* (Suppl. 1): 133-138.

Kazanji M, Heraud JM, Merien F, Pique C, Thé G, Gessain A, Jacobson S 2006 Chimeric peptide vaccine composed of B- and T-cell epitopes of human T-cell leukemia virus type 1 induces humoral and cellular immune responses and reduces the proviral load in immunized squirrel monkeys (*Saimiri sciureus*). *J Gen Virol 87*: 1331-1337.

Korber B 1997. HIV Signature and sequence variation analysis. In AG Rodrigo, GH Learn *Computational Analysis of HIV Molecular Sequences*, Kluwer Academic Publisher, Dordrecht, Netherlands, p. 55-72.

Kuroki M, Nakamura M, Itoyama Y, Tanaka Y, Shiraki H, Baba E, Esaki T, Tatsumoto T, Nagafuchi S, Nakano S. 1992. Identification of new epitopes recognized by human monoclonal antibodies with neutralizing and antibody dependent cellular cytotoxicity activities specific for human T cell leukemia virus type 1. *J Immunol 149*: 940-948.

Kyte J, Doolittle RF 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*: 105-132.

Mueller N 1991. The epidemiology of HTLV-I infection. *Cancer Causes Control 2*: 37-52.

Nagy K, Clapham P, Cheingsong-Popov R, Weiss RA 1983. Human T-cell leukemia virus type I: induction of syncytia and inhibition by patients' sera. *Int J Cancer 32*: 321-328.

Nicholas KB, Nicholas HBJ, Deerfield DW 1997. GeneDoc: Analysis and visualization of genetic variation. *EMB News 14*: 30.

Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M 1986. HTLV-I associated myelopathy, a new clinical entity. *Lancet 1*: 1031-1032

Palker TJ, Tanner ME, Scearce RM, Streilein RD, Clark ME, Haynes BF 1989. Mapping of immunogenic regions of human T cell leukemia virus type I (HTLV-I) gp46 and gp21 envelope glycoproteins with env-encoded synthetic peptides and a monoclonal antibody to gp46. *J Immunol 142*: 971-978.

Parker JM, Guo D, Hodges RS 1986. 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 25*: 5425-5432.

Parker CE, Daenke S, Nightingale S, Bangham CR 1992. Activated, HTLV-I-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. *Virology 188*: 628-636.

Pique C, Connan F, Levilain JP, Choppin J, Dokhélar MC 1996. Among all human T-cell leukemia virus type 1 proteins, tax, polymerase, and envelope proteins are predicted as preferential targets for the HLA-A2 restricted cytotoxic T-cell response. *J Virol 70*: 4919-4926.

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci 77*: 7415-7419.

Queiroz AT, Mota-Miranda AC, de Oliveira T, Moreau DR, Urpia C de C, Carvalho CM, Galvão-Castro B, Alcântara LC 2007. Re-mapping the molecular features of the human immunodeficiency virus type 1 and human T-cell lymphotropic virus type 1 Brazilian sequences using a bioinformatics unit established in Salvador, Bahia, Brazil, to give support to the viral epidemiology studies. *Mem Inst Oswaldo Cruz 102*: 133-139.

Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics 50*: 213-219.

Schönbach C, Nokihara K, Bangham CR, Kariyone A, Karaki S, Shida H, Takatsu K, Egawa K, Wiesmüller KH, Takiguchi M 1996. Identification of HTLV-1-specific CTL directed against synthetic and naturally processed peptides in HLA-B* 3501 transgenic mice. *Virology 226*: 102-112.

Seiki M, Hattori S, Hirayama Y, Yoshida M 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci 80*: 3618-3622.

Sundaram R, Lynch MP, Rawale SV, Sun Y, Kazanji M, Kaumaya PT 2004. De novo design of peptide immunogens that mimic the coiled coil region of human T-cell leukemia virus type-1 glycoprotein 21 transmembrane subunit for induction of native protein reactive neutralizing antibodies. *J Biol Chem 279*: 24141-24151.

Tallet B, Astier-Gin T, Moynet D, Londos-Gagliardi D, Guillemain B 2001. Sequence variations in the amino- and carboxy-terminal parts of the surface envelope glycoprotein of HTLV type 1 induce specific neutralizing antibodies. *AIDS Res Hum Retroviruses 17*: 337-348.

Tanaka Y, Zeng L, Shiraki H, Shida H, Tozawa H 1991. Identification of a neutralization epitope on the envelope gp46 antigen of human T cell leukemia virus type I and induction of neutralizing antibody by peptide immunization. *J Immunol 147*: 354-360.

Yang Z 1997. PAML: a program package for phylogenetic analysis by maximum likelihood . *Comput Appl Biosci 13*: 555-556.

Yang Z, Nielsen R, Goldman N, Pedersen AM 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics 155*: 431-449.