

RESEARCH

Open Access

Molecular characterization of HTLV-1 gp46 glycoprotein from health carriers and HAM/TSP infected individuals

Aline C A Mota-Miranda^{1,2,3}, Fernanda K Barreto¹, Maria F C Amarante⁵, Everton Batista⁴, Joana P Monteiro-Cunha^{2,3}, Lourdes Farre⁴, Bernardo Galvão-Castro^{1,2} and Luiz C J Alcantara^{1*}

Abstract

Background: Human T-cell Leukemia Virus type 1 (HTLV-1) is the etiological agent of tropical spastic paraparesis/HTLV-associated myelopathy (HAM/TSP) that can be identified in around 0.25%–3.8% of the infected population. Disease progression can be monitored by the proviral load and may depend on genetic factors, however, it is not well understood why some HTLV-1 infected people develop the disease while others do not. The present study attempts to assess the molecular diversity of gp46 glycoprotein in HAM/TSP patients and Health Carrier (HC) individuals.

Methods: Blood samples were collected from 10 individuals, and DNA was extracted from PBMCs to measure the HTLV-1 proviral load. The gp46 coding sequences were amplified PCR, cloned and sequenced. The molecular characterization was performed using bioinformatics tools.

Results: The median HTLV-1 proviral load of HC (n = 5) and HAM/TSP (n = 5) patients was similar (average 316,227 copies/10⁶ PBMCs). The gp46 molecular characterization of 146 clones (70 HC and 76 HAM/TSP) revealed an overall diversity, within HC and HAM/TSP clones, of 0.4% and 0.6%, respectively. Five frequent mutations were detected among groups (HAM/TSP and HC clone sequences). A single amino acid (aa) substitution (S35L) was exclusive for the HC group, and three gp46 substitutions (F14S, N42H, G72S) were exclusive for the HAM/TSP group. The remaining frequent mutation (V247I) was present in both groups (p = 0.0014). The in silico protein analysis revealed that the mutated alleles F14S and N42H represent more hydrophilic and flexible protein domains that are likely to be less antigenic. The Receptor Binding Domain is quite variable in the HAM/TSP group. Two other domains (aa 53–75 and 175–209) that contain multiple linear T-cell epitopes showed genetic diversity in both HAM/TSP and HC groups. Further analysis revealed 27 and 13 T-cell epitopes for class I HLA alleles and class II HLA alleles, when analyzing the entire gp46.

Conclusions: The most common gp46 mutations were not associated clinical status because they were found in only one individual, except for the V247I mutation, that was found at viral clones from HAM/TSP and HC individuals. Because of this, we cannot associate any of the gp46 found mutations with the clinical profile.

Keywords: HTLV-1, HAM/TSP, Gp46, Mutation

* Correspondence: lalcan@bahia.fiocruz.br

¹Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, CPqGM/
FIOCRUZ, Rua Waldemar Falcão 121, Brotas, Salvador, Bahia 40295-001, Brazil
Full list of author information is available at the end of the article

Background

HTLV-1 was initially isolated from a patient with cutaneous T-cell lymphoma in early 1980s, being the first discovered human retrovirus [1]. HTLV-1 is classically the etiological agent of tropical spastic paraparesis/HTLV-associated myelopathy (HAM/TSP), which is characterized, besides the spastic paraparesis, by urinary disturbances and, in almost all patients, spasticity and/or hyper-reflexia of the lower extremities [2]. HTLV-1 has also been implicated as the cause of Adult T-cell Leukemia/Lymphoma (ATLL) and of other inflammatory disorders such as arthritis, uveitis, dermatitis, lymphadenitis and Sjogren's syndrome [2].

It is estimated that around 0.25%–3.8% of the HTLV-1 infected population develops HAM/TSP [3-6]. Despite the scarcity of data on the real prevalence of HAM/TSP in the Brazilian population, the individual risk of disease development is of considerable importance for HTLV-1 endemic areas such as southern Japan, the Caribbean, Central and South America (including Brazil), the Middle East, Melanesia, and equatorial regions of Africa [4]. The progression to HAM/TSP includes viral factors such as proviral load and host genetic factors such as Human Leukocyte Antigen (HLA). Most clinical studies associate proviral load to increased risk of disease progression, and in some of them, the HTLV-1 proviral load progressively increases according to HAM/TSP ascertainment level (possible, probable, and definite) of the De Castro-Costa diagnostic criteria [7,8]. On the other side, molecular data from southern Japan indicates that the alleles HLA-A*02 and Cw*08 (HLA class I) are associated with a significant reduction in both HTLV-1 proviral load and the risk of HAM/TSP, whereas HLA-B*5401 (HLA class I) and HLA-DRB1*0101 (HLA class II) predispose to HAM/TSP in the same population [9,10].

Although most studies of HTLV-1 genotype have reported no association between HTLV-1 variants and risk to HAM/TSP, it is known that HAM/TSP patients generally have higher anti-HTLV-1 antibody titer than Health Carriers (HC) with a similar proviral load [11]. Additionally, the antibody reactivity to Tax epitopes can be different between HAM/TSP (71%–93%) and HC (27%–37%) individuals [12]. Regarding the regulatory protein genotypes, a p12 K88 allele was found mainly among HAM/TSP individuals, suggesting that it could be implicated as a pathogenesis marker [13].

The HTLV-1 envelope glycoproteins (Env) mediate the binding of the virus to its receptor on the surface of target cells and the subsequent fusion of virus and cell membranes. It is also important to note that three different molecules are assumed to participate in HTLV-1 entry: a glucose transporter type 1 (GLUT-1), the VEGF-165 receptor Neuropilin 1 (NRP-1), and heparan sulfate proteoglycans (HSPG) [14-16]. Throughout the characterization of neutralizing antibodies and analysis of Env mutants, it is

now known that gp46 contains a great number of functional and immunodominant domains that are associated to the prevalence of linear epitopes, interaction to neuropilin 1 or heparan sulfate molecules, and induction of antibodies [13,17-20].

Because it is not well understood why some HTLV-1 infected people develop disease while others do not, the purpose of the present study was to assess the molecular diversity of gp46 glycoprotein in HAM/TSP and HC individuals to identify the potential molecular biomarkers and/or the therapeutic targets for HAM/TSP. Despite the lack of association with the clinical status, five frequent mutations at gp46 were identified.

Results

Sample characterization

Blood samples were collected from 10 HTLV-1 infected individuals recruited from a reference health unit: 5 HC and 5 HAM/TSP patients. Because the progression to HAM/TSP is multifactorial and is not only dependent on viral factors, but also on the host's genetic background, we choose all individuals with similar proviral load and age. The median HTLV-1 proviral load of HC and HAM/TSP patients was the same (316,227 copies/10⁶ PBMCs) (Fisher Test, $p = 0.0007$ -after Bonferroni Correction) (Figure 1), while the median age was 53 (38–72) and 58 (41–73) years for HC and HAM/TSP patients, respectively ($p = 0.239$).

Molecular characterization of gp46 glycoprotein

We obtained 146 gp46 clone sequences (908pb, 302aa): 70 from HC (average of 14 clones per individual) and 76 from HAM/TSP individuals (average of 15 clones per individual). The HC and HAM/TSP clone sequence overall diversity was 0.4% and 0.6%, respectively. Regarding the amino acid changes, we identified five mutations with

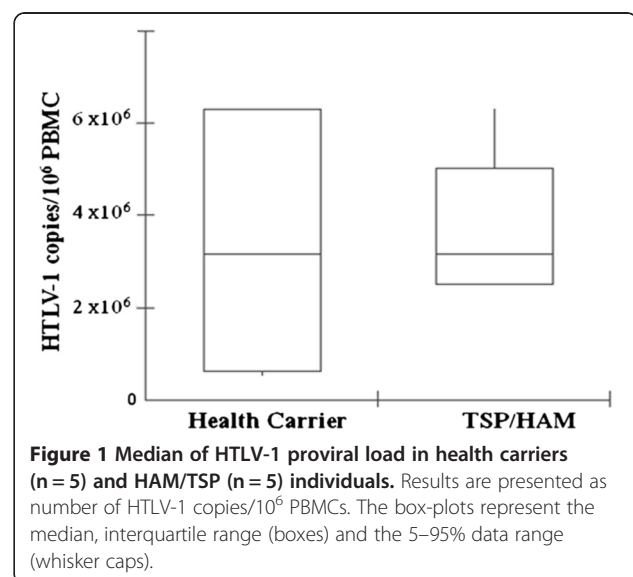
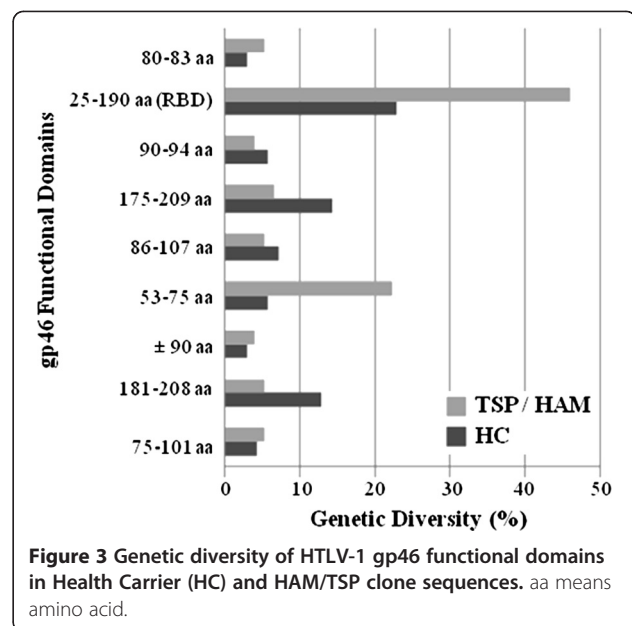
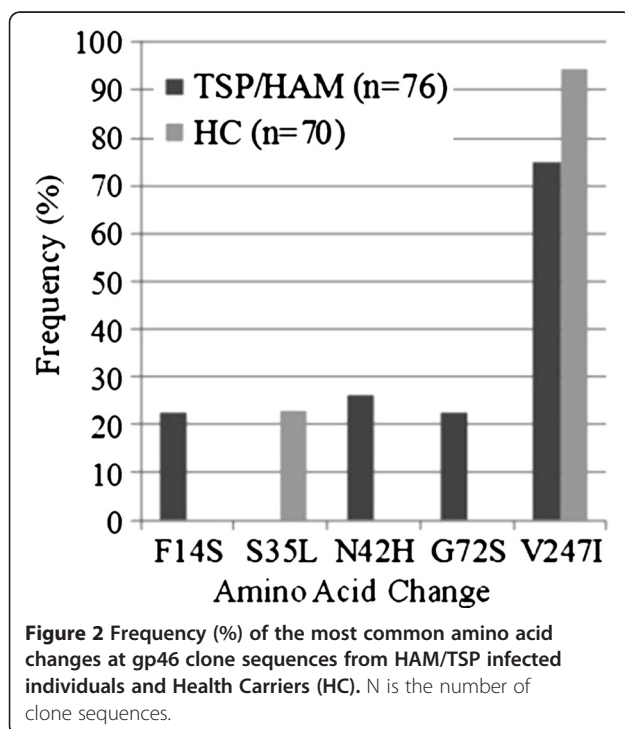


Figure 1 Median of HTLV-1 proviral load in health carriers (n = 5) and HAM/TSP (n = 5) individuals. Results are presented as number of HTLV-1 copies/10⁶ PBMCs. The box-plots represent the median, interquartile range (boxes) and the 5–95% data range (whisker caps).

frequency over 20% within the clone groups (Figure 2). Among these common mutations, one (S35L) was exclusive to the HC group, three (F14S, N42H, G72S) were exclusive to the HAM/TSP group and, finally, one mutation (V247I) was found in both groups but with a statistically significant difference in the frequency within the groups (Fisher Test, $p = 0.0014$).

Among the five mutations found, three (F14S, S35L and N42H) were not present in any of the 42 Brazilian sequences used to create the consensus. G72S was found in one *Brazilian* sequence; however, we do not have information about the geographic region from which the sequence was obtained. Finally, the V247I mutation was found in sequences from all different geographic regions represented in our dataset. Similarly, F14S, S35L, N42H and V247I mutations were not present in any sequence of the *Cosmopolita dataset*, while G72S mutation was found in sequences from Gabon, Martinique, French Guyana and Guadalupe.

We used the already known gp46 functional domains to perform the search for mutations (See Additional file 1). In general, the Receptor Binding Domain (RBD) presented the greatest number of amino acid changes in both groups. However, the general diversity (frequent and not frequent mutations) profile is different between the HC and HAM/TSP groups. The HC group presented the three most divergent domains, and one of them is also a divergent domain in the HAM/TSP group (Figure 3). It is interesting to note that the second most divergent domain (175-209 aa) in the HC group is characterized by the



prevalence of linear epitopes, and it was not so divergent in the HAM/TSP group. It was also noticeable that the domain 53-75 aa, associated to the most prevalent occurrence of linear epitopes, had the second higher index of divergence among the HAM/TSP group (Figure 3). This domain is less divergent in the HC group of clones. The 197 amino acid site, which was already associated to the decrease in cellular fusion, was mutated (D197N) in two clone sequences in our data set: one from the HC group and the other from the HAM/TSP group (data not shown). Moreover, one HAM/TSP sequence was mutated (Y80C) at position 80 aa, which is part of the YSLY motif involved in the protein trafficking to the plasma membrane.

We also performed the epitope prediction in the domains with the higher genetic diversity in HAM/TSP (53-75 aa) and HC clones (175-209 aa). Analyzing the 175-209 aa domain, 16 HLA-I and seven HLA-II specific epitopes were found (Table 1). On the other hand, 11 HLA-I and five HLA-II specific epitopes were identified in the 53-75 aa domain (Table 2). Since the G72S mutation is located in the 53-75 aa domain, we tested both wild type and mutated alleles. The mutated allele was associated with change in HLA binding score in three of the 11 epitopes including this position (72 aa). For the HLA A26 and A01 (HLA-I) specific epitopes, starting at positions 65 aa and 68 aa, the HLA binding score increased and decreased, respectively (Table 2). For the HLA DRB1 0101 (HLA-II) specific epitope, starting at position 64 aa, the HLA binding score decreased (Table 2).

To identify if the amino acid mutations found has been positively selected, we have performed positive selective pressure analysis. According to our results, the ω value calculated for the clone sequences from HAM/

Table 1 Possible epitopes for classes I and II HLA alleles in HTLV-1 gp46 175-209 aa domain

A-Class I HLA alleles																	
AA position	Epitope sequence															HLA allele	
175	F	L	N	T	E	P	S	Q	L								A 0201, B 08
177	N	T	E	P	S	Q	L	P	P								A 01, A 1101, B 4501
179	E	P	S	Q	L	P	P	T	A								A 0702
182	Q	L	P	P	T	A	P	P	L								A 0201, B 08, B 4402
183	L	P	P	T	A	P	P	L	L								A 0701, B 08, B 5101
184	P	P	T	A	P	P	L	L	P								B 0702
185	P	T	A	P	P	L	L	P	H								A 1101
188	P	P	L	L	P	H	S	N	L								B 0702, B 02, B 1402, B 2705, B 37, B 5101
191	L	P	H	S	N	L	D	H	I								B 4901, B 5101
192	P	H	S	N	L	D	H	I	L								B 1510, B 3801 B 3901
193	H	S	N	L	D	H	I	L	E								A 1101
194	S	N	L	D	H	I	L	E	P								B 4501
199	I	L	E	P	S	I	P	W	K								A 03, A 1101, A 6801, B 2705
202	P	S	I	P	W	K	S	K	L								A 26, B 2705
203	S	I	P	W	K	S	K	L	L								A 0201, B 08, B 37
205	P	W	K	S	K	L	L	T	L								A 26, B 3901
B-Class II HLA alleles																	
AA position	Epitope sequence															HLA allele	
172	P	I	W	F	L	N	T	E	P	S	Q	L	P	P	T	DRB 0701	
180	P	S	Q	L	P	P	I	A	P	P	L	L	P	H	S	DRB 0301, DRB 0701	
187	A	P	P	L	L	P	H	S	N	L	D	H	I	L	E	DRB 0301, DRB 1501	
188	P	P	L	L	P	H	S	N	L	D	H	I	L	E	P	DRB 0701	
193	H	S	N	L	D	H	I	L	E	P	S	I	P	W	K	DRB 0701	
194	S	N	L	D	H	I	L	E	P	S	I	P	W	K	S	DRB1 0101	
201	E	P	S	I	P	W	K	S	K	L	L	T	L	V	Q	DRB 1501	

TSP individuals was higher than the ω value calculated for the clone sequences from HC individuals. Despite this result no calculated ω value was >1 .

The identification of post-translation modification sites is summarized in Table 3 and shows that the same potential protein domains were found among the HC and HAM/TSP groups of sequences. However, despite being the same sites, their frequencies were statistically different, being more frequent among the viral clones isolated from HAM/TSP individuals (Table 3).

We submitted wild type and mutated sequences to physico-chemical analysis using the NPSA tool. The following was observed: 1) the mutated allele presenting the F14S mutation was more hydrophilic and flexible than the wild type allele (Figure 4A); 2) the mutated allele presenting the N42H mutation was less antigenic than the wild type allele (Figure 4B); 3) the mutated allele presenting the V247I mutation was less antigenic than the wild type allele (Figure 4C). According to our results, other less frequent mutations were associated to changes in the physico-chemical profile and potential protein domains: the T142I mutation was associated to

the loss of a N-glycosylation site and to a decrease in antigenicity; the F159S mutation presented a more hydrophilic and flexible profile, associated to the insertion of a N-Myristilation site; and the H290D, C289R and F141S mutations were characterized by an increase in hydrophilicity and flexibility, the opposite of the D197N mutation which was associated to a decrease in flexibility (Data not shown).

After submission of wild type and mutated sequences to SWISS-MODEL server, the results showed that the S35L and G72S mutations were associated to the change of coil for extended beta at the secondary structure prediction (Figure 5).

Discussion

The low overall genetic diversity found in gp46 sequences is in agreement with the fact that the HTLV-1 genome exhibits relatively few sequence variations, emphasizing that recombinant envelope glycoproteins or synthetic peptides could be used as an effective vaccine candidate. This genomic stability and the presence of neutralizing antibodies in HTLV-1-infected individuals

Table 2 Possible epitopes for classes I and II HLA alleles in HTLV-1 gp46 53-75 aa domain

A-Class I HLA alleles																
AA position	Epitope sequence													HLA allele		
52	L	D	L	L	A	L	S	A	D						B 37	
53	D	L	L	A	L	S	A	D	Q						A 03	
54	L	L	A	L	S	A	D	Q	A						A 0201, A 03	
55	L	A	L	S	A	D	Q	A	L						A 0201, B 08, B 1402, B 5101	
56	A	L	S	A	D	Q	A	L	Q						A 03	
57	L	S	A	D	Q	A	L	Q	P						A 1101	
62	A	L	Q	P	P	C	P	N	L						A 0702, B 08, B 2705, B4402	
63	L	Q	P	P	C	P	N	L	V						B 3902	
64	Q	P	P	C	P	N	L	V	G						B 0702, B 5101	
65	P	P	C	P	N	L	V	G	Y						A 01, A 26, B 4402	
68	P	N	L	V	G	Y	S	S	Y						A 01	
B-Class II HLA alleles																
AA position	Epitope sequence													HLA allele		
51	T	L	D	L	L	A	L	S	A	D	Q	A	L	Q	P	DRB1 0101, DRB 0401
52	L	D	L	L	A	L	S	A	D	Q	A	L	Q	P	P	DRB 0301, DRB 0701
54	L	L	A	L	S	A	D	Q	A	L	Q	P	P	C	P	DRB1 0101, DRB 07071, DRB 1501
64	Q	P	P	C	P	N	L	V	G	Y	S	S	Y	H	A	DRB1 0101
67	C	P	N	L	V	G	Y	S	S	Y	H	A	T	Y	S	DRB1 0101, DRB 0401, DRB 0701, DRB 1501

are favorable factors for the design of an efficient HTLV-1 vaccine, and this could be an advantage comparing to the challenge of developing a vaccine to HIV infection. However, it is important to note that the sequences from HAM/TSP individuals are relatively more divergent compared to those from HC individuals, confirming that the immune environment represents effective selective pressure.

It is also worth noting that the relative lack of variation in the HTLV gp46 relative to the HIV gp120 is believed to reflect the fact that HTLV-1 persists in an individual by clonal expansion of HTLV-1-infected cells, rather than continual spread of the virus, as in HIV infection. However, there are minor sequence variations

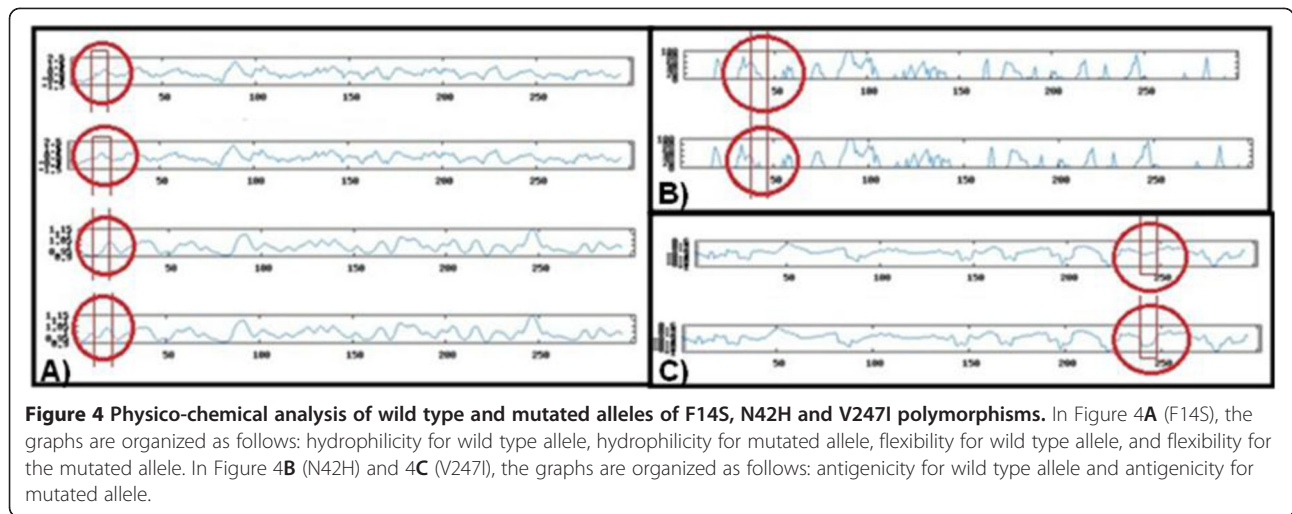
between geographical regions and certain HTLV-1 subgroups that can be associated with different risks of HAM/TSP. Furthermore, a quarter of the provirus load may be carried by CD8+ T cells and when equilibrium is reached between a persistently replicating pathogen and the immune response, the frequency of specific CD8+ T cells is an unreliable index of the efficiency or effectiveness of the T cell response [21].

We identified five common mutations in all 176 gp46 sequences. Four of the most common mutations (except the V247I mutation) were associated to an intra-host profile because were specifically found in the same single individual: the S35L mutation was exclusive for one HC individual, the F14S and G72S mutations were exclusive

Table 3 Potential protein domains in HTLV-1 gp46 from health carriers and HAM/TSP individuals

Potential protein domain	Health carrier (N = 70)		HAM/TSP (N = 76)	
	Amino acid site	Frequency (%)	Amino acid site	Frequency (%)
N-Myristilation site	97-102	91.4	97-102	100*
CK2-phosphorylation site	103-106	91.4	103-106	98.7 [§]
	194-197	88.6	194-197	100 [¶]
N-glycosilation	140-143	90	140-143	98.7 ^⓪
	222-225	91.4	222-225	100 [†]
	244-247	91.4	244-247	98.7 [‡]
	272-275	92.8	272-275	98.7 [#]
PKC-phosphorylation site	109-111	91.4	109-111	100 ^Δ

The numbers superscript at the last column represents the results from the Exact Fisher Test to the frequency of the potential protein domains in the two different groups. * $p = 0.028$, [§] $p = 0.0551$, [¶] $p = 0.0022$, ^⓪ $p = 0.0283$, [†] $p = 0.0108$, [‡] $p = 0.0551$, [#] $p = 0.1047$, ^Δ $p = 0.0108$. N means the number of clones.

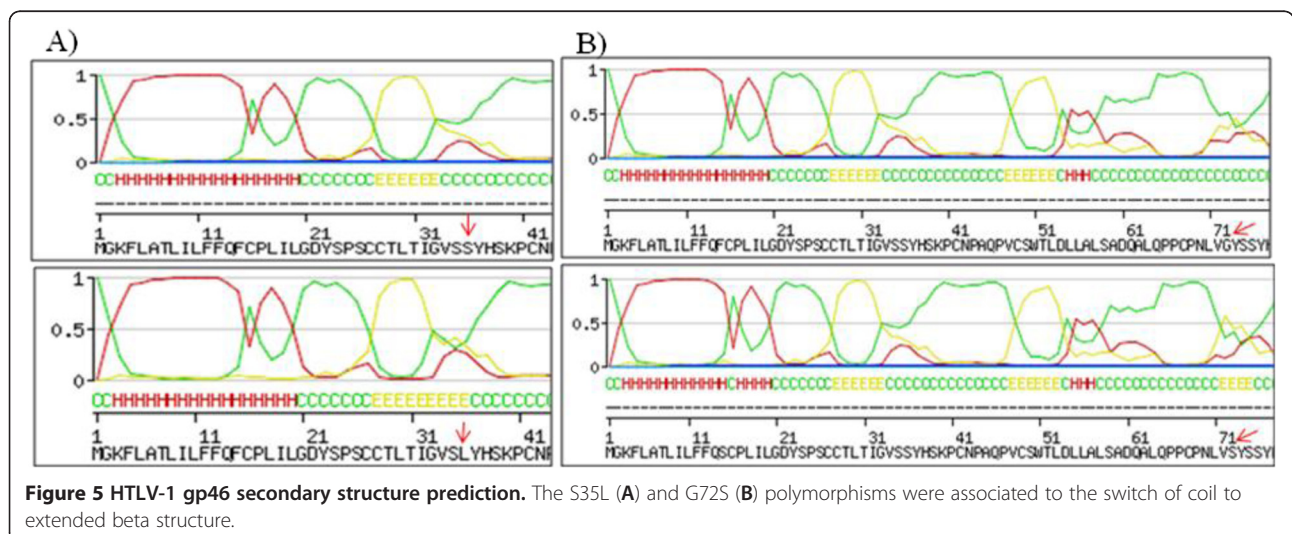


for one HAM/TSP individual, and finally the N42H change was exclusive for other HAM/TSP individual. Because of this, we cannot associate any of these mutations to the clinical profile, due to the reduced number of individuals recruited in this study. Due to the observed intra-host profile of the found mutations, we compared them to the proviral load of each individual. The mutations S35L, F14L, G72S and N42H were identified in individuals with high proviral load ($\log > 5$).

We suggest that future studies evaluating the molecular diversity of envelope proteins should take into account a larger number of individuals, and a reduced number of clones. A larger number of individuals could be useful to evaluate the association of the possible mutations to the clinical status, while cloning could be useful for the identification of intra-host variations and possible emerging changes. In this regard, three (F14S, S35L and N42H) of these newly identified mutations can be classified as a completely new protein characteristics, while the V247I

mutation is an already known characteristic of Brazilian HTLV-1 isolates.

We could identify that HTLV-1 gp46 from HC and HAM/TSP individuals had different genetic diversity profiles, especially in domains with prevalence of linear epitopes. In these domains, we identified different epitopes between HC and HAM/TSP individuals. Recent data revealed that patients with HAM/TSP recognized a significantly broader repertoire of Env epitopes, compared with health carriers, suggesting that the diversity, frequency and repertoire of HTLV-1 Env CD8⁺ T-cells may be related to the hyperimmune response in HAM/TSP disease [22]. Additionally, the identification of the Env epitopes associated with the immune system activation could be useful to vaccine design [23]. In a recent report, the immunogenicity and protective efficacy in squirrel monkeys of a B-cell env gene epitope and of a T-cell multi-epitope derived from Tri-Tax were tested. The authors showed that the reduction of the number of



cells producing IFN- γ in response to the Tri-Tax T-cell epitope peptide was much lower than with the B-cell env gene epitope peptide, indicating the involvement of other virus-specific T-cell subsets in this cellular immune response [24]. These T-cell subsets could be those ones that are specific to divergent viruses.

We believe in the potential of these mutations to become clinical markers as they are located in important protein domains. The fact that some of these mutations have been associated to physico-chemical or secondary structure changes does not mean that there is no need for *in vitro* assays. This investigation is important because antibodies are most likely to recognize linear epitopes but they are also dependent on the conformation of the antigen and efficient epitope recognition determines the neutralizing properties of anti-HTLV-1 antibodies [25]. In this case, non effective antibodies could be closely related to the clinical manifestation because they can play an important role in controlling HTLV-1 dissemination and the viral burden on natural infections, in particular through antibody-dependent cellular cytotoxicity [26].

The genomic regions where the post-translational sites were identified, showed a high identity among the sequences, suggesting the influence of these sites on the host immune response, and virus latency. The post-translational modifications in envelope glycoproteins are not well documented therefore there is a lack of evidence of the importance of these sites for virus clonal expansion. However, these sites are frequently found in HTLV-1 envelope proteins, and it would be interesting to investigate the functional impact of them. Regarding the Tax protein, for example, much is known about the post-translational modifications and they are important in the constitutive activation of NF- κ B pathways, inhibition of DNA repair, activation of p53 tumor suppressor and cell cycle control [27].

The immune responses that differentiate health carrier from HAM/TSP could be a result of the fact that infected cells, which display the envelope on the surface, remain insensitive and not vulnerable to antibody-dependent cellular cytotoxicity and antibody-mediated complement fixation. From the present results, our group is carrying out a site-directed mutagenesis assay to test the impact of the described mutations in the binding capacity of the variant protein against natural antibodies produced by HC and HAM/TSP individuals. The benefits of investigating these mutations can also be useful for the Western Blot indeterminate profile which is dependent on the efficient binding of gp46 and natural antibodies.

Conclusions

In conclusion, the results presented in this study suggest that the most frequent gp46 mutations (F14S, N42H, V247I and S35L), as well as, the mutations (H290D,

C289R, F141S, D197N and F159S) associated to the loss of a post-translational site should be evaluated in greater detail, in particular the reactivity of the mutated proteins against natural antibodies. Furthermore, we believe that analyses involving a greater number of sequences can provide more information about the occurrence of these variations and the identification of others, which are just as important as the ones presented here.

Methods

Patients

Blood samples were collected from 10 individuals who were frequently seen at the Bahia School of Medicine and Public Health, HTLV reference center in Salvador, located in Northeastern Brazil, in the first period of 2010. This is a free public outpatient clinic that has provided comprehensive care to a total of 1,050 patients since 2002. All 10 patients were included in the present study with inclusion criteria consisting of an available HTLV-1 proviral load measurement and a neurological evaluation. All patients had a quite similar proviral load and median of age. All cases of co-infection were excluded, as well as symptomatic patients with other HTLV-1-associated diseases, such as infective dermatitis, uveitis, ATLL, Sicca syndrome, etc. The Disability Status Scale (DSS) [28] and the Osame Motor Disability Score (OMDS) were regularly applied by a neurologist. This study was approved by the Ethical Committee (Protocol Number 81/2007) of the Bahia School of Medicine and Public Health. Informed consent was obtained from all enrolled patients. Salvador is the capital of Bahia State in northeast Brazil and presents wide socioeconomic differences. The population is roughly 80% black or racially mixed African and Portuguese descendents.

Measurement of HTLV-1 proviral load

PBMCs were obtained from EDTA blood by density gradient centrifugation. DNA was extracted using spin column DNA extraction system (Qiagen, Hilden, Germany) and HTLV-1 proviral load was quantified using a real-time TaqMan Polymerase Chain Reaction (PCR) method, as described previously [29]. Albumin DNA was used as an endogenous reference. The value of HTLV-1 proviral load was reported as the [(HTLV-1 average copy number)/(albumin average copy number)] $\times 2 \times 10^6$ and expressed as the number of HTLV-1 copies per 10^6 cells in PBMCs.

gp46 PCR and cloning

The gp46 coding sequences from the viruses isolated from the PBMCs of HTLV-1-infected patients were amplified by PCR. In brief, gp46F/gp46R (gp46F 5'CGC CGATCCCAAAGAAAAA3'/gp46R 5'ACATGGAGCCG GTAATCCC3') were designed and used to amplify a 936pb fragment, corresponding to the entire gp46 coding

region. The PCR products were purified using the Qiaquick Gel Extraction kit (Qiagen, QIAamp DNA miniKit, Qiagen, Hilden, Düsseldorf, Germany). The amplicons were then ligated into the pCR[®]4-TOPO available at TOPO TA cloning[®] kit for sequencing (Invitrogen, CA, USA), and the mini preparations were made using the kit (Wizard[®] Plus Minipreps DNA Purification System). All purified clones were sequenced in an ABI Prism 3100 DNA Sequencer (Applied Biosystems Inc., Foster City, CA) using Taq FS Dye (Applied Biosystems) terminator cycle sequencing. Besides the same PCR primers, a new inner primer (gp46 R2 5'GACGTGCCAAGTGGATAGGC3') was used to optimize the sequencing reactions.

gp46 Molecular characterization

Before analyzing the obtained gp46 sequences for the presence of mutations and identification of molecular characteristics, two datasets were constructed. The first was called "*Cosmopolita database*" and was composed by 27 gp46 sequences from different countries (except from Brazil), previously available in the NCBI/Nucleotide Sequence Database (GenBank). It is important to note that the 27 sequences of the "*Cosmopolita dataset*" did not have any information about subtype and were generated from different countries: French Guyana, Central America (n = 7); Martinique, Central America (n = 1); Gabon, Central Africa (n = 12); and Guadalupe, Central America (n = 7). The second dataset was called "*Brazilian dataset*" and was composed by 42 gp46 sequences from Brazil, previously available in GenBank. It is therefore important to note that these 42 Brazilian sequences are classified as subtype a (*Cosmopolita*), and were generated from different Brazilian geographic regions: Salvador, Northern (n = 21); Londrina, South (n = 5); São Paulo, Southeast (n = 11), while five of them did not present any information about their geographic origin.

Both datasets were submitted separately to the Clustal X software [30] to perform the alignment which was then manually edited using the GeneDoc program [31], and finally, the edited alignments were used to generate a unique consensus sequence of each dataset using Bioedit software [32]. The consensus sequence from "*Cosmopolita database*" was called "*Cosmopolita reference*" and the consensus sequence from the "*Brazilian dataset*" was called "*Brazilian reference*". These consensus sequences comprise the most frequent nucleotide variants found in previously published gp46 sequences from Brazil and elsewhere. These consensus sequences were used as the reference sequences to identify possible mutations in the 146 newly generated gp46 sequences.

The genetic distances were measured within the two distinct groups: gp46 sequences from HC and HAM/TSP HTLV-1 infected individuals. The Tamura Nei model was used with a distance matrix implemented in the MEGA

3.0 package [33], and the standard error computation was obtained by Bootstrap analysis (1000 replicates). The mutation/polymorphism identification was performed manually using the visualization of alignment in the Bioedit software.

To test the hypothesis that the amino acid substitutions within the gp46 sequences could have been favored or not by natural selection, the positive selection was assessed using six different codon-based maximum-likelihood substitution models [34]. All models were implemented in the HYPHY program [35] and 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 labeled 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 positively selected sites [23,36].

Epitope prediction was carried out for the HC and HAM/TSP consensus sequences to 27 HLA-I (HLA A1101, HLA A26, HLA B1510, HLA B4402, HLA A01, HLA A0201, HLA A2402, HLA A5101, HLA A03, HLA B2705, HLA A6801, HLA B08, HLA B0702, HLA B2709, HLA B 1402, HLA B1501, HLA B18, HLA B37, HLA B3801, HLAB 3901, HLA B 3902, HLA B4001, HLA B4101, HLA B 4501, HLA B 4701, HLA B 4901, HLA B5101) 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 (<http://www.syfpeithi.de/Scripts/MHCServer.dll/Epitope Prediction.htm>) [37]. 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.

To investigate the possible influence of the described mutations in the gp46 sequences, physico-chemical analysis was performed using Network Protein Sequence Analysis (NPSA) (<http://npsa-pbil.ibcp.fr/>) [38-42] and the potential protein domain analysis using the GeneDoc software and the Prosite tool, as previously described [43].

Finally, the SWISS-MODEL online tool (<http://swissmodel.expasy.org/>) [44] was used as a fully automated protein structure homology-modeling server, to infer the possible influence of the amino acid changes at protein secondary structure.

All *env* nucleotide sequences previously deposited in the GenBank and used in the study are listed below with their corresponding accession number: [L26585, L26586, L33265, L33266, AF091494-AF091500, AF092065, L76041-L76049, L76052-L76054, L76056, L76058, L76060, DQ007189-DQ007209, HM770426-HM770440, U81865-U81869, AF077209].

Additional file

Additional file 1: Brief description of previously published domains in the HTLV-1 gp46 protein and their exact amino acid location.

Competing interest

The authors declare they have no competing interests.

Authors' contributions

ACAMM carried out the molecular genetic studies, the bioinformatics analyses and drafted the manuscript. FKB participated in the molecular genetic studies and the bioinformatics analyses. MFCA participated in the cloning. EB participated in the cloning and molecular studies. JPMC had performed the English review and helped to draft the manuscript. LfV contributed to the analysis of the results and helped to draft the manuscript. BGCF contributed with the blood samples and clinical data from HTLV-1 infected individuals. LCJA conceived of the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the FAPESB and CNPq financial support.

Author details

¹Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, CPqGM/FIOCRUZ, Rua Waldemar Falcão 121, Brotas, Salvador, Bahia 40295-001, Brazil.

²Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia, Brazil.

³Universidade Federal da Bahia, Instituto de Ciências da Saúde, Salvador, Bahia, Brazil. ⁴Laboratório de Patologia Experimental, Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil. ⁵National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

Received: 12 September 2012 Accepted: 27 February 2013

Published: 6 March 2013

References

- Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC: **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 USA* 1980, **77**:7415–7419.
- Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M: **HTLV-I associated myelopathy, a new clinical entity.** *Lancet* 1986, **1**:1031–1032.
- Proietti FA, Carneiro-Proietti ABF, Catalan-Soares B, Murphy EL: **Global epidemiology of HTLV-I infection and associated diseases.** *Oncogene* 2005, **24**:6058–6068.
- Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, De Thé G: **Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis.** *Lancet* 1985, **2**:407–410.
- Kaplan JE, Osame M, Kubota H, Igata A, Nishitani H, Maeda Y, Khabbaz RF, Janssen RS: **The risk of development of HTLV-I-associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I.** *J Acquir Immune Defic Syndr* 1990, **3**:1096–1101.
- Maloney EM, Cleghorn FR, Morgan OS, Rodgers-Johnson P, Cranston B, Jack N, Blattner WA, Bartholomew C, Manns A: **Incidence of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in Jamaica and Trinidad.** *J Acquir Immune Defic Syndr Hum Retrovirology* 1998, **17**:167–170.
- Nagai M, Usuku K, Matsumoto W, Kodama D, Takenouchi N, Moritoyo T, Has Hashiguchi S, Ichinose M, Bangham CR, Izumo S, Osame M: **Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP.** *J Neurovirology* 1998, **4**:586–593.
- Grassi MFR, Olavarria VN, Kruschewsky RA, Mascarenhas E, Dourado I, Correia LCL, Castro-Costa CM, Galvão-Castro B: **Human T cell lymphotropic virus type 1 (HTLV-1) proviral load of HTLV-associated myelopathy/tropical spastic paraparesis (HAM/TSP) patients according to New diagnostic criteria of HAM/TSP.** *J Medical Virology* 2011, **83**:1269–1274.
- Jeffery KJM, Usuku K, Hall SE, Matsumoto W, Taylor GP, Procter J, Bunce M, Ogg GS, Welsh KI, Weber JN, Lloyd AL, Nowak MA, Nagai M, Kodama D, Izumo S, Osame M, Bangham CR: **HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy.** *Proc Natl Acad Sci USA* 1999, **96**:3848–3853.
- Jeffery KJM, Siddiqui AA, Bunce M, Lloyd AL, Vine AM, Witkover AV, Izumo S, Usuku K, Welsh KI, Osame M, Bangham CRM: **The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection.** *J Immunology* 2000, **165**:7278–7284.
- Ishihara S, Okayama A, Stuver S, Horinouchi H, Shioiri S, Murai K, Kubota T, Yamashita R, Tachibana N, Tsubouchi H, Mueller N: **Association of HTLV-I antibody profile of asymptomatic carriers with proviral DNA levels of peripheral blood mononuclear cells.** *J Acquir Immune Defic Syndr* 1994, **7**:199–203.
- Lal RB, Giam CZ, Coligan JE, Rudolph DL: **Differential immune responsiveness to the immunodominant epitopes of regulatory proteins (tax and rex) in human T cell lymphotropic virus type I-associated myelopathy.** *J Infectious Diseases* 1994, **169**:496–503.
- Iniguez AM, Otsuki K, Magalhães GP, Silva EA, Vicente ACP: **Genetic markers on the HTLV-1 p12I protein sequences from Brazilian HAM/TSP patients and asymptomatic HTLV-1 carrier isolates.** *AIDS Res Human Retroviruses* 2005, **21**(6):580–582.
- Manel N, Kim FJ, Kinet S, Taylor N, Sitbon M, Battini JL: **The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV.** *Cell* 2003, **115**:449–459.
- Ghez D, Lepelletier Y, Lambert S, Fourneau JM, Blot V, Janvier S, Arnulf B, van Enderet PM, Heveker N, Pique C: **Neuropilin-1 is involved in human T-cell lymphotropic virus type 1 entry.** *J Virol* 2006, **80**:6844–6854.
- Jones KS, Petrow-Sadowski C, Bertolette DC, Huang Y, Ruscetti FW: **Heparan sulfate proteoglycans mediate attachment and entry of human T-cell leukemia virus type 1 virions into CD4+ T cells.** *J Virol* 2005, **79**:12692–12702.
- Seki M, Sashiyama H, Hayami M, Shida H: **Intracellular processing and immunogenicity of the envelope proteins of human T-cell leukemia virus type I that are expressed from recombinant vaccinia viruses.** *Virus genes* 1990, **3**:235–249.
- Sagara Y, Inoue Y, Shiraki H, Jinno A, Hoshino H, Maeda Y: **Identification and mapping of functional domains on human T-cell lymphotropic virus type 1 envelope proteins by using synthetic peptides.** *J Virol* 1996, **70**(3):1564–1569.
- Rosenberg A, Delamarre L, Preira A, Dokhelar MC: **Analysis of functional conservation in the surface and transmembrane glycoprotein subunits of human T-cell leukemia virus type 1 (HTLV-1) and HTLV-2.** *J Virol* 1998, **72**(9):7609–7614.
- Ilinskaya A, Heidecker G, Derse D: **Opposing effects of a tyrosine-based sorting motif and a PDZ-binding motif regulate human T-lymphotropic virus type 1 envelope trafficking.** *J Virol* 2010, **84**:6995–7004.
- Bangham CRM: **The immune control and cell-to-cell spread of human T-lymphotropic virus type 1.** *J Gen Virol* 2003, **84**:3177–3189.
- Kosako T, Akimoto M, Toji S, Tomohiro K, White Y, Suzuki S, Arima T, Suruga Y, Matsushita K, Shimeno H, Soeda S, Kubota R, Izumo S, Uozumi K, Arima N: **Target epitopes of HTLV-1 recognized by class I MHC-restricted cytotoxic T lymphocytes in patients with myelopathy and spastic paraparesis and infected patients with autoimmune disorders.** *J Med Virol* 2011, **83**:501–509.
- Mota-Miranda AC, de Oliveira T, Moreau DR, Bomfim C, Galvão-Castro B, Alcântara LCJ: **Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design.** *Mem Inst Oswaldo Cruz* 2007, **102**(6):741–749.
- Kazanji M, Heraud JM, Merien F, Pique C, de Thé G, Gessain A, Jacobson S: **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* 2006, **87**:1331–1337.
- Ghez D, Lepelletier Y, Jones KS, Pique C, Hermine O: **Current concept regarding the HTLV-1 receptor complex.** *Retrovirology* 2010, **7**:99–110.
- Kuo CS, Mirsalotis A, Brighty DW: **Antibodies to the envelope glycoprotein of human T cell leukemia virus type I robustly activate cell-mediated cytotoxic responses and directly neutralize viral infectivity at multiple steps of the entry process.** *J Immunology* 2011, **187**:361–371.
- Shembade N, Harhaj ED: **Role of post-translational modifications of HTLV-1 Tax in NF- κ B activation.** *World J Biol Chem* 2010, **26**:13–20.
- Kurtzke JF: **A new scale for evaluating disability in multiple sclerosis.** *Neurology* 1955, **5**:580–583.

29. Dehee A, Cesaire R, Desire N, Lezin A, Bourdonne O, Bera O, Plumelle Y, Smadja D, Nicolas JC: **Quantification of HTLV-I proviral load by a TaqMan real-time PCR assay.** *J Virol Methods* 2002, **102**:37–51.
30. Jeanmougin F, Thompson JD, Goy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with clustal X.** *Trends Biochem Sci* 1998, **23**:403–405.
31. Nicholas KB, Nicholas HBJ, Deerfield DW: **GeneDoc: analysis and visualization of genetic variation.** *Embnew News* 1997, **4**:14.
32. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41**:95–98.
33. Kumar S, Tamura K, Nei M: **MEGA: molecular evolutionary genetics analysis software for microcomputers.** *Cabios* 1994, **10**:189–191.
34. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–449.
35. Pond SLK, Muse SP: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2000, **21**:676–679.
36. De Oliveira T, Salemi M, Gordon M, Vandamme AM, Van Rensburg EJ, Engelbrecht S, Coovadia HM, Cassol S: **Mapping sites of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design?** *Genetics* 2004, **167**:1047–1058.
37. Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanovic S: **SYFPEITHI: database for MHC ligands and peptide motifs.** *Immunogenetics* 1999, **50**:213–219.
38. Argos P, Rao JK, Hargrave PA: **Structural prediction of membrane-bound proteins.** *Eur J Biochem* 1982, **15**:565–575.
39. Kyte J, Doolittle R: **A simple method for displaying the hydropathic character of a protein.** *J Mol Biol* 1982, **157**:105–132.
40. Hopp TP, Woods KR: **Physico-chemical profiles: a computer program for predicting protein antigenic determinants.** *Mol Immunol* 1983, **20**:483–489.
41. Karplus PA, Schulz GE: **Prediction of chain flexibility in proteins - a tool for the selection of peptide antigens.** *Naturwissenschaften* 1985, **72**:212–213.
42. Parker JMR, Guo D, Hodges RS: **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**:5425–5431.
43. Queiroz AT, Mota-Miranda AC, De Oliveira T, Moreau DR, Urpia CDE, Carvalho CM, Galvão-Castro B, Alcantara LCJ: **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* 2007, **102**:133–139.
44. Arnold K, Bordoli L, Kopp J, Schwede T: **The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling.** *Bioinformatics* 2006, **22**:195–201.

doi:10.1186/1743-422X-10-75

Cite this article as: Mota-Miranda et al.: Molecular characterization of HTLV-1 gp46 glycoprotein from health carriers and HAM/TSP infected individuals. *Virology Journal* 2013 **10**:75.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

