**ESCOLA BAHIANA DE MEDICINA E SAÚDE PÚBLICA**
**CURSO BIOMEDICINA**

**MARINA SILVEIRA CUCCO**

# CARACTERIZAÇÃO MOLECULAR DAS REGIÕES GENÔMICAS hbz E LTR DO HTLV-1 EM PACIENTES COM DIFERENTES MANIFESTAÇÕES CLÍNICAS

**SALVADOR – BA**
**2020**

**MARINA SILVEIRA CUCCO**

# CARACTERIZAÇÃO MOLECULAR DAS REGIÕES GENÔMICAS hbz E LTR DO HTLV-1 EM PACIENTES COM DIFERENTES MANIFESTAÇÕES CLÍNICAS

Trabalho de Conclusão de Curso apresentado à Escola Bahiana de Medicina e Saúde Pública, como parte dos requisitos para obtenção do título de Bacharel em Biomedicina.

Orientador: Profª. Drª. Luciane Amorim Santos

**SALVADOR – BA**

**2020**

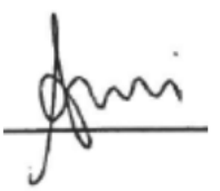
**MARINA SILVEIRA CUCCO**

**CARACTERIZAÇÃO MOLECULAR DAS REGIÕES GENÔMICAS hbz E LTR DO HTLV-1 EM PACIENTES COM DIFERENTES MANIFESTAÇÕES CLÍNICAS**

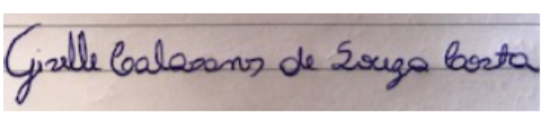
Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do grau de Bacharel em Biomedicina e aprovada em sua forma final pelo Curso de Biomedicina da Escola Bahiana de Medicina e Saúde Pública.

Salvador – BA, 25 de junho de 2020.

Profª. Drª. Luciane Amorim Santos
Escola Bahiana de Medicina e Saúde Pública



Profª. Drª. Fernanda Khouri Barreto
Universidade Federal da Bahia



Profª. Drª. Giselle Calasans de Souza Costa
Universidade Federal da Bahia

*A minha família dedico este trabalho pois a crença deles no meu sucesso fez toda a diferença.*

## AGRADECIMENTOS

Ao universo agradeço pelos caminhos apontados e pelas pessoas trazidas para perto.

Aos meus pais e minha irmã por sempre terem me apoiado, desde criança, incentivando meus estudos e me presenteando com as melhores oportunidades. Tudo sempre foi e sempre será por vocês e para vocês, pois sem vocês eu nada seria.

Ao resto da minha família, que mesmo de longe sempre se fez presente na minha vida e sempre se mostrou orgulhosa de mim, até mesmo quando eu não me sentia merecedora. Vocês são o meu maior tesouro.

A todos os meus grupos de amigos por sempre me colocarem para cima, partilharem da minha caminhada, sempre aumentando minha autoestima e me fazendo seguir em frente, vocês são parte de quem eu sou.

Ao grupo de pesquisa da Bahiana e da Fiocruz, por todo apoio técnico, emocional e por todo conhecimento compartilhado, meus sinceros agradecimentos.

A minha orientadora, que me foi presenteada desde o início da minha caminhada acadêmica, agradeço imensamente por ter confiado em mim e ter me proporcionado tamanho aprendizado e oportunidade.

E por fim, agradeço a mim que, apesar de muito ter pensado em desistir, me mantive forte e persistente para terminar com êxito a minha caminhada da graduação.

*“Sai de casa sempre assim que der*
*Mas sai sem esquecer que a sua casa é sempre aqui.”*
*Outro Eu*

## RESUMO

O Vírus Linfotrópico de Células T Humanas do tipo 1 (HTLV-1) afeta cerca de 5 a 10 milhões de pessoas no mundo, que podem se manter assintomáticas (AC) ou evoluírem para doenças inflamatórias e/ou patologias severas, como paraparesia espástica tropical/mielopatia associada ao HTLV-1 (HAM/TSP). O genoma viral é flanqueado por regiões LTR e contém genes estruturais, genes responsáveis pelas enzimas virais e uma região pX que codifica proteínas acessórias e regulatórias, a exemplo de Tax, p12 e HBZ. O objetivo deste estudo é descrever a caracterização molecular das regiões de hbz e LTR a partir de amostras de pacientes infectados pelo HTLV-1. Foram selecionadas 15 amostras de pacientes AC e HAM/TSP do CHTLV da Escola Bahiana de Medicina e Saúde Pública. Foi realizada PCR de todas as amostras, os produtos foram checados através de gel de agarose e, os que amplificaram pelo menos uma região de interesse, foram purificados e sequenciados (Sanger). Além disso, foram selecionadas 571 sequências de hbz e 151 de LTR publicadas nos bancos de dados Genbank e HTLV-1 Molecular Epidemiology Database. O programa Geneious Prime, foi utilizado para alinhamento, identificação das mutações nas duas regiões de interesse e mapeamento de sítios de transcrição (TFBS) de LTR. Para a região de *hbz*, o perfil físico-químico das sequências e sítios de modificação pós-traducionais foram analisados. A genotipagem das sequências foi definida através de árvore filogenética, usando a região LTR. Neste artigo, são descritas duas mutações não-sinônimas de HBZ, V17M e R121Q, encontradas em frequências diferentes nas sequências de AC e HAM/TSP quando comparados. Essas mutações alteraram propriedades físico-químicas e podem ter um papel no estabelecimento da sintomatologia. Dentre os TFBS descritos, alterações nos sítios Sp1 foram encontrados tanto no CHTLV quanto nos bancos de dados e podem ser importantes na expressão de HBZ. Este trabalho identifica possíveis fatores de diferenciação das manifestações clínicas entre diferentes pacientes infectados por HTLV-1, como as mutações V17M e R121Q e a exclusão de TFBS por conta de mutações, a exemplo da G418A.



**Palavras-chave**: caracterização molecular, hbz, LTR.

**SUMÁRIO**

## 1. Artigo Científico

# MOLECULAR CHARACTERIZATION OF HTLV-1 GENOMIC REGIONS *hbz* AND LTR FROM PATIENTS WITH DIFFERENT CLINICAL OUTCOMES

Cucco M S[1], De-Moraes L E P[2,3], Khouri R[2,3], Galvão-Castro B[1], Araujo T H[1], Barreto F K[4], Santos L A[1,2].

[1]Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia, Brasil.

[2]Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brasil.

[3]Universidade Federal da Bahia, Salvador, Bahia, Brasil.

[4]Universidade Federal da Bahia, Vitoria da Conquista, Bahia, Brasil.

Running title: HBZ and LTR molecular characterization

Key words: HTLV-1, molecular characterization, HBZ, LTR, mutation, TFBS

**Abstract**

Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) affects from 5 to 10 million people around the world, which can remain asymptomatic (AC) or evolve to inflammatory disease and severe pathologies as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). The virus genome is flanked by LTR regions and contains structural genes, genes responsible for coding viral enzymes and pX region, which is responsible for coding regulatory proteins like Tax, p12, HBZ. In this study we aim to describe the molecular characterization of *hbz* and LTR regions from HTLV-1 infected patients. Fifteen samples from AC and HAM/TSP patients available in HTLV Center (CHTLV) from Bahia School of Medicine and Public Health. PCR technique was used with all samples, amplicons were checked in an agarose gel and the samples that amplified at least one interest regions (*hbz* and/or LTR) were purified and sent to Sanger sequencing. In addition, 571 *hbz* and 151 LTR sequences were selected from Genbank e HTLV-1 Molecular Epidemiology Database. Geneious prime was used to perform alignment, variant calling and map transcription factor binding sites (TFBS). Physicochemical profile and post translation sites from *hbz* mutated sequences were analyzed. Genotyping was performed through phylogenetic tree using LTR sequences. Here is described two non-synonymous mutations in HBZ protein, V17M and R121Q, that were found in different frequencies in AC sequences when compared to HAM/TSP. These mutations alter some of the protein physicochemical characteristics and may be important to clinical status establishment. From the described TFBS, alterations in Sp1 binding sites, found in sequences from CHTVL and databse, can be an important factor to HBZ expression. This study identifies possible factors involved in the clinical status differentiation in HTLV-1 patients, as V17M and R121Q mutation and the abrogation of TFBS caused by LTR mutation as G418A.

## Introduction

The Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) infection was first observed in Japan in 1977 in patients with Adult T-Cell Leukemia/Lymphoma, known as ATLL[1]. But the virus was only isolated and described in 1980 – being the first human retrovirus described – from an American T-Cell Cutaneous Lymphoma patient[2].

HTLV-1 is a worldwide distributed pathogen affecting from 5 to 10 million people[3]. People infected with HTLV-1, despite not knowing why, can remain asymptomatic (AC) or can evolve to different symptomatology like Infective dermatitis associated with the HTLV-1 (IDH), HTLV-1-associated myelopathy/tropical spastic paraparesis known as HAM/TSP (0.25-3.8% of patients) and ATLL (2-5% of patients)[4]. It is estimated that there are 2.5 million people are infected with HTLV-1 in Brazil, which characterizes the country as one with the highest absolute number of cases worldwide[5]. North and Northeast regions of Brazil are the most affected by HTLV-1[3],with emphasis to the city of Salvador, capital of Bahia[6].

The HTLV-1 genome is composed by structural genes as *gag*, *pol*, *env* and a non-structure region known as pX, flanked by two regions known as long terminal repeat (LTR). The pX region, close to 3' LTR, encodes regulatory and accessory proteins as p12, p8, p30, Tax, Rex and HBZ[7].

The LTR regions are divided in 3 subregions: U3, R e U5, whose sequences contain information that are extremely important to the genetic material reverse transcription, replication and genome integration[8]. Genome integration can occur at any site of the host genome and may cause interruption and inactivation of several host genes. The LTR sequences can also participate in the coordination of viral genes expression after the genome is integrated[9]. It has been detected that LTR mutations can be related to the increase of proviral load and possible development of HAM/TSP[10].

Proviral load (pVL) correspond to the amount of peripheral blood mononuclear cells (PBMC) that have HTLV-1 provirus integrated at the host genome[11]. It has been described the pVL it may be a marker for the disease progression in the patients, being the higher pVL correlated with HAM/TSP and ATLL development, but the studies are controversial[12–14].

HTLV-1 bzip domain (HBZ), named after the bzip domain of the protein, was first described in 2002 and it is encoded by the minus-strand RNA that is transcribed in the antisense direction by a promoter present in the 3' LTR[15]. For what is known, HBZ is a regulatory protein that can induce proliferation of infected cells and inhibit apoptosis[16], cooperating in cell immortalization, maintenance and multiplication[17]. Another important regulatory protein is Tax, responsible for inducing viral proliferation and increase of genetic instability[18]. Also, HBZ and Tax function are related to expression regulation of accessory proteins[10] but HBZ antagonizes many Tax functions as the LTR trans-activation, negatively regulating this action[6,19,20].

Besides, it was also found an isoform from HBZ protein, the HBZ-SI, generated by alternative splicing. Both have >95% similarity at the amino acid sequence, which can also imply in similar functions, and their difference is the N termini. Similar expression levels between the two forms of the protein were observed in ATL cells[21]. HBZ is constantly expressed by HTLV-1 infected cells[22] and therefore, it can maybe be used as a new target for therapeutic approach. HBZ cell localization appears to be exclusive in the cytoplasm in HAM/TSP patients while in ATLL individuals is seems to appear like dots in the nucleus as well as in the cytoplasm. It might be possible that HBZ sequence variation may interfere in the protein localization[23].

Therefore, the present study aims to evaluate genetic variation of HTLV-1 *hbz* and LTR genomic regions from sequences originated from patients with four different symptomologies and whether the modifications influence the different clinical outcomes.

**Materials and Methods**

Study population

Samples were collected from HTLV-1 infected patients – AC and HAM/TSP – followed up at Centro de Atendimento ao Portador de HTLV (CHTLV) of Escola Bahiana de Medicina e Saúde Pública. It is important to highlight that the asymptomatic patients have mild symptoms that does not allow to classify them to a clinical status, but they do feel the impact of the infection in the physical and psychological health. Clinical status of all HTLV-1 patients was defined by the CHTLV neurologists following WHO criteria. Selected patients were invited to participate of the study and signed the informed consent form. Clinical and epidemiological data were obtained from the filed medical records at CHTLV. This study was approved by the Research Ethics Committee of Escola Bahiana de Medicina e Saúde Pública (approval number 464.286).

Sample collecting and processing

Blood samples were collected at CHTLV and transferred to Gonçalo Moniz Institute (IGM) where the peripheral blood mononuclear cells (PBMC) separation was performed at the same day. DNA from PBMC was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following manufactory instructions.

Polymerase Chain Reaction (PCR) and sequencing

The hbz and LTR regions were amplified through a PCR cycling, using Veriti Thermal Cycler (Applied Biosystems, United States), according to protocol previously published[24] using the primers 24+ (5'-CGTATCGCCTCCCTCGCGCCATCAGAGTATGCTGCCC AGAACAG-3') and 27- (5'-CTATGCGCCTTGCCAGCCCGCTCAGGGTTCCATGTATC CATTTCGGA-3') for hbz gene, and primers 30+ (5'-CGTATCGCCTCCCTCGCGCCATC AGCCAGCCATCTTTAGTACTACAGT-3') and 32- (5'-CTATGCGCCTTGCCAGCCCGC TCAGAGCCAACGGAG TCGC-3') for 3' LTR. Once the amplification of the interest region was confirmed by agarose gel electrophoresis, PCR products were purified using QIAquick

PCR Purification Kit (Qiagen, Hilden, Germany) and sequenced using a BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, United States) with capillary electrophoresis run on an 3500xL Genetic Analyzer (Applied Biosystems, Foster City, California, United States) employing the same primers as used in amplification. Raw Sanger sequences files were used to assembly a consensus sequence for each sample, generating the CHTLV dataset. Sequence visualization, editing and assembly were performed using Geneious Prime 2020.1.1 (https://www.geneious.com).

Database sequence collection and dataset assembly

In order to characterize mutations of HTLV-1 hbz and LTR regions and compare the frequency of mutations between different clinical status from CHTLV dataset and data already published, we build a dataset from sequences available at Genbank (https://www.ncbi.nlm.nih.gov/nuccore) and HTLV-1 Molecular Epidemiology Database (https://htlv1db.bahia.fiocruz.br). Sequence visualization, editing and assembly were performed using Geneious Prime 2020.1.1 (https://www.geneious.com).

Genbank sequence search was performed using “HTLV-1 AND hbz/p8/p12/p30”, “HTLV-1 AND LTR” and “HTLV-1 AND complete genome” as key words. For the HTLV-1 Molecular Epidemiology Database sequences search the regions of interest, LTR and hbz, were selected on the criteria box. After retrieving available data, sequences isolated from species other than human, sequences from different types of HTLV, duplicated sequences and those without clinical information were excluded and the remaining sequences were aligned with ATK1 (J02029.1) as reference. After the alignment, sequences that were in our genome regions of interest (hbz and LTR) were included.

HTLV-1 Genotyping

To identify the HTLV-1 Subtypes and Subgroup, phylogenetic analyses were performed with CHTLV and database LTR sequences. The sequences from each dataset were aligned to reference sequences (AF054627, AY342303-AY342304, AY818425, DQ005552-DQ005555, DQ005564-DQ005567, EF672333-EF672336, FJ853491, GU225731-GU225732, J02029, L02534, L76309, L76312, M37299, U12804-U12806, Y16476-Y16477, Y16479-Y16482, Y16484, Y16487 and Y17014) from all HTLV-1 subtypes and subgroups using MAFTT v. 1. 4. 0 (Katoh, Standley 2013) implemented at Geneious Prime 2020.1.1 (https://www.geneious.com). Then, Maximum Likelihood (ML) trees were reconstructed using IQ-TREE v.1.6.12 (MINH et al., 2020) under TN+F+G4 nucleotide substitution modeling, inferred by ModelFinder implemented in the IQ-TREE software (KALYAANAMOORTHY et al., 2017). Ultrafast bootstrap analysis with 1000-replicate (HOANG et al., 2018) was used to calculate the statistical support of the tree branches. All trees were visualized and edit using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Genetic analysis of hbz and LTR

Geneious Prime annotation tool was used to find variations in both CHTLV and database datasets and TFScan plugin was used to map transcription factor binding sites at the LTR region. The found mutations and the transcriptional factors binding sites were then compared between CHTLV dataset and database dataset.

Physical-chemical profile and the search for post-translational modification against the PROSITE database was performed using PRABI platform (https://npsa-prabi.ibcp.fr/) to analyze whether HBZ mutations alter the protein characteristics comparing to the wild-type (ATK1). The physical-chemical profiles were plotted using GraphPad Prism 8.4.2 (GraphPad Software, San Diego, California, USA).

Statistical analyses

To detect if there were any differences in the mutation frequency in the different clinical groups, fisher's exact test was performed using GraphPad Prism 5.02 (GraphPad Software, San Diego, California, USA). The sequences were divided in the following groups: AC; HAM/TSP, ATLL and IDH. The statistical differences between the observed frequencies of the mutations was performed by pairs using AC as control against each clinical outcome (HAM/TSP, ATLL and IDH). A p-value $< 0.05$ was considered as statistically significant.

## Results

Samples and sequences from CHTLV

Samples from 28 patients were included in order to perform PCR (hbz and LTR) and 15 samples have been successfully amplified for at least one region of interest. We obtained 15 sequences of hbz region and 14 sequences of LTR region. Sequencing quality of LTR region was not the same in all sequences, resulting in eight sequences with low coverage (145pb-248pb) and six sequences with medium/high coverage (481pb-651pb). Out of this, 13 were from HAM/TSP patients and two were AC.

*hbz* molecular characterization

All sequences from CHTLV successfully aligned with ATK1 hbz region. After aligning and performing variant calling a total of 27 mutations were detected at the 15 hbz sequences generated in this study, all of them classified as single nucleotide variants (SNVs). Six mutations are present in 100.0% of sequences, ten are present in frequencies above 10.0% and below 90.0% and 11 are present in only one sequence. Among the SNVs, 19 were synonymous mutations and eight are non-synonymous mutations, resulting in an amino acid change, one change resulting in a stop codon.

A total of 571 hbz sequences from database were included in the study after applying all the selection criteria and then aligned with ATK1 in the selected genome region. A total of 352 mutations were detected in the database dataset, 129 synonymous mutations and 223 non-synonymous mutations, after the variant calling. From the 27 CHTLV dataset mutations, 12 were present in the database dataset. Mutation frequency comparison between the CHTLV dataset and the database dataset is showed in Table 1.

To evaluate if any detected mutation were related to the clinical profile, CHTLV dataset was combined with de database dataset and all sequences were divided in groups according to

clinical status: 211 were from AC patients, 297 from HAM/TSP, 67 from ATL and 11 from IDH.

A total of 13 common mutations were detected between the clinical groups, 136 mutations presented in at least 2 of the clinical groups, 97 mutations exclusives for HAM/TSP, 103 exclusives for AC, 16 exclusives for ATL and 2 exclusives for IDH. Most mutations found were not frequent among the sequences or did not present an amino acid change.

To identify if there were any differences in the mutation frequencies between the HAM/TSP, ATLL and IDH groups when compared to the AC group, the fisher exact statistical analysis was performed in all non-synonymous mutations in frequencies above 10,0%. Two mutations, V17M and R121Q, showed a difference in the mutation frequency in the comparison between AC and HAM/TSP groups, with statistical support ($p<0,0001$ and $p=0,0005$ respectively). The V17M was only detected at HAM/TSP group and this could indicate that it can possibly be a signature from this clinical group. For the IDH and ATL groups no difference in the frequency of the mutations were detected with statistical support.

The non-synonymous SNV R121Q were found in 53,3% of CHTLV dataset and 19,4% of database dataset. To understand the possible effect of this amino acid change on the protein structure, physicochemical analysis of HBZ was performed. According to the profile, the modifications in comparison to the wild-type are the reduction of the protein hydrophilicity, at the accessibility and antigenicity A at the mutation site, while hydropathy, flexibility, membrane buried-helix and antigenicity B have no changes (Figure. 1).

Non-synonymous mutation V17M was found only in database HAM/TSP group with 16.3% of frequency. This mutation appears to be responsible for a slight reduction of hydropathy and antigenicity B, while the other characteristics remains the same as the wild-type (Figure. 2).

When searched for post-transduction modification sites in HBZ, both mutations described showed no changes and neither created any site.

LTR molecular characterization

To avoid the subtype/subgroup classification bias, LTR sequences were used to identify the subtype of sequences from both datasets by phylogenetics analyses. All CHTLV sequences were classified as Cosmopolitan Transcontinental (HTLV-1 aA) (Figure 3) and database sequences were divided in other subtypes/subgroups besides HTLV-1 aA (data not shown). Only the database sequences from Cosmopolitan Transcontinental (HTLV-1 aA) were included in the study, totalizing 151 LTR database sequences.

After aligning and performing variant calling at the LTR CTHLV dataset, nine SNVs, one deletion and one insertion were found. In the database dataset, a total of 202 mutations were detected in the LTR region, being 169 mutations in frequencies below 10,0% or above 90,0%. Comparing mutations from both datasets, nine of 11 mutations found in CHTLV dataset were also found in the database dataset.

In order to find if any mutations in the CHTLV dataset were responsible for changing the transcriptional factors binding sites (TFBS), creating or abrogating sites, both CHTLV and database datasets were screened to identify the TFBS. Eight CHTLV sequences were excluded from this analysis due to low coverage. In total 32 TFBS were mapped, being four new added TFBS and five were absent due to presence of mutations (Table 2). One of the mutations (G418A) was present in a Sp1 binding site and its presence abrogated the Sp1 site and created a new ER binding site.

The same mapping for TFBS was performed to LTR database dataset and a total of 184 TFBS were found. From this total, 35 were new added TFBS and 25 TFBS were absent also due to mutation. No similar change between CHTLV TFBS and database TFBS were found.

**Discussion**

HTLV-1 infection in Brazil affects approximately 2.5 million individuals [3] being one of the epicenters in the world. It is important to remember that most infected individuals remain AC during the course of infections[4] and that this study aims to contribute in the identification of possible reasons why some of the patients evolve to a serious clinical status, as HAM/TSP. All data generated by CHTLV dataset were compared to database dataset to describe similarities between the two datasets.

In order to achieve this purpose, HTLV-1 genomic regions from HAM/TSP and AC samples were sequenced to identify HBZ protein mutations and TFBS at LTR region. From all HBZ non-synonymous mutations, two were found with different frequency between AC and HAM/TSP groups with statistical support, V17M and R121Q. Also, one mutation at LTR region (G418A) that changed TFBS and could affect the viral gene expression was detected[20].

One mutation found in both CHTLV and database dataset, R121Q, is present in sequences from all clinical groups included in this study but is most frequent in AC group. Statistical analysis of mutation frequencies demonstrated that when comparing each group (HAM/TSP, ATL and IDH) to the AC group, only the comparison between AC and HAM/TSP were statistically significant while the other did not have statistical support.

The other mutation described in our HBZ analysis, V17M, appears to be a signature from HAM/TSP patients since it was not found in the AC group neither in sequences from the other clinical status included in this study and it has not been described yet. HBZ is an antisense protein, originated from reverse reading frame, and is homologous to the sense reading frame. So mutations found in this region can possibly interfere in the other proteins functions from pX gene[25].

Both mutations resulted in some physicochemical alteration. Amino acid change V17M have altered hydropathy and antigenicity in the region where is located in the protein, slightly reducing these two properties. The other amino acid change, R121Q, alters hydrophilicity, accessibility and antigenicity, also reducing these properties in the region where is located. Since there was not found any reports for these mutations, a detailed analysis to understand if these alterations affect the protein function is needed.

Another important scenario is HBZ-Tax correlation, as the two proteins do not interact directly to one another, but HBZ can interfere Tax-dependent viral transcription through inhibition of the binding between Tax and transcription sites, as CREB and TxRE III, found to be mediated through the bzip domain from HBZ[15]. However, even the truncated form from HBZ lacking bzip domain it is still effective in down-regulating viral transcription, which suggests that other domain from HBZ is also capable of accomplish this effect[26]. So, it may be possible that mutations found in this region alter this function of HBZ.

Beyond the function of down-regulating the viral transcription[15], HBZ has another function as inhibiting apoptosis through suppression of *Bim*[27], which is responsible for inducing apoptosis (intrinsic pathway), located in the cytoplasmatic membranes[28]. The suppression of *bim* gene occurs due to the inhibition of exportation of FoxO3a from the nucleus to the cytoplasm and so it cannot bind to the forkhead responsive element (FHRE) at *bim* promoter[27]. Therefore, the mutations are to be investigated whether they affect this HBZ function.

In agreement to what is shown by Dissinger et al (2014), that HBZ has post-translational modifications sites (PTMs), such as acetylation and phosphorylation ones, but the sites observed in the CHTLV dataset did not changed nor created any new sites. Also, HBZ has more

acetylation PTM once it interacts with acetyltransferases. Besides this, little is known about the PTMs and its function[29].

The genotyping analysis from CHTLV dataset identified that all sequences are subtype Cosmopolitan and subgroup Transcontinental (aA), as it shown at ML tree, agreeing with previously studies[30,31]. That confirms that in Brazil this is the most common subtype/subgroup circulating. One of the reasons of the importance of subtyping is to exclude subtype-specific alterations that could give false results in comparing mutations and TFBS[10].

HBZ transcription is dependent on the 3' LTR and Sp1 binding site is an important TFBS to this process[20]. Mapping these sites is essential to analyze whether the protein is being expressed and its function are in normal conditions. In this study we found an abrogation of one Sp1 binding site in one sequence of CHTLV dataset caused by the G418A mutation, which changed a Sp1 to an ER binding site (Table 2). Missing Sp1 sites were also detected in the database dataset. A thorough analysis is needed to comprehend how this alteration can affect HBZ expression.

All the findings in this study are to be validated once more sequences are generated and analyzed. It is important to increase the number of LTR sequences and reach a similar sample number from two principal groups compared (AC and HAM/TSP). One of the possible explanation for most our hbz sequences being from HAM/TSP patients is that hbz expression may not occur in AC or if it is expressed it is in too low levels[32]. Furthermore, functional tests comparing wild-type and mutated protein may be a way to elucidate if mutations found in the sequences studied altered any protein functions.

## Conclusion

This study identified two possible mutations in hbz region with amino acid change, V17M and R121Q, that can imply in HBZ protein function and also possibly change other proteins functions. Besides this, the mutations can possibly be used as a prognostic factor for the development of HAM/TSP, once the mutations showed statistical support when comparing AC and HAM/TSP, being one exclusive to the HAM/TSP sequences. In addition, abrogation of transcription factors binding sites were observed, as the Sp1 abrogation in some LTR sequences, and it may be important to genes expression, as hbz.

**Funding Information**

This research was supported by Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

**References**


1. Uchiyama T, Yodoi J, Sagawa K, Takatsuki K, Uchino H. Adult T-cell leukemia: clinical and hematologic features of 16 cases. Blood [Internet]. 1977 Sep;50(3):481–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/301762
2. 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 U S A. 1980;77(12 II):7415–9.
3. Gessain A, Cassar O. Epidemiological aspects and world distribution of HTLV-1 infection. Front Microbiol. 2012;3(NOV):1–23.
4. Yamano Y, Sato T. Clinical pathophysiology of human T-lymphotropic virus-type 1-associated myelopathy/tropical spastic paraparesis. Front Microbiol. 2012;3(NOV):1–10.
5. Carneiro-Proietti ABF, Ramos Ribas JG, Catalan-Soares BC, Martins ML, Brito-Melo GEA, Martins-Filho OA, et al. Infecção e doença pelos vírus linfotrópicos humanos de células T (HTLV-I/II) no Brasil. Rev Soc Bras Med Trop. 2002;35(5):499–508.
6. Catalan-Soares B, Carneiro-Proietti AB de F, Proietti FA. Heterogeneous geographic distribution of human T-cell lymphotropic viruses I and II (HTLV-I/II): serological screening prevalence rates in blood donors from large urban areas in Brazil. Cad saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Esc Nac Saúde Pública. 2005;21(3):926–31.
7. Giam CZ, Semmes OJ. HTLV-1 infection and adult T-cell leukemia/ lymphoma-A tale of two proteins: Tax and HBZ. Viruses. 2016;8(6).
8. Grant C, Barmak K, Alefantis T, Yao J, Jacobson S, Wigdahl B. Human T cell leukemia virus type I and neurologic disease: Events in bone marrow, peripheral blood, and central nervous system during normal immune surveillance and neuroinflammation. J Cell Physiol. 2002;190(2):133–59.
9. Varmus HE. Form and function of retroviral proviruses. Science (80- ).

1982;216(4548):812–20.

10. Rego FFDA, De Oliveira T, Giovanetti M, Galvão-Castro B, De Souza Gonçalves M, Alcantara LC. Deep sequencing analysis of Human T cell lymphotropic virus Type 1 long terminal repeat 5’ region from patients with tropical spastic Paraparesis/Human T cell lymphotropic virus Type 1-associated myelopathy and asymptomatic carriers. AIDS Res Hum Retroviruses. 2016;32(3):279–83.
11. Pineda MV, Bouzas MB, Remesar M, Fridman A, Remondegui C, Mammana L, et al. Relevance of HTLV-1 proviral load in asymptomatic and symptomatic patients living in endemic and non-endemic areas of Argentina. PLoS One. 2019;14(11):1–10.
12. Demontis MA, Hilburn S, Taylor GP. Human T cell lymphotropic virus type 1 viral load variability and long-term trends in asymptomatic carriers and in patients with human T cell lymphotropic virus type 1-related diseases. AIDS Res Hum Retroviruses. 2013;29(2):359–64.
13. Okayama A, Stuver S, Matsuoka M, Ishizaki J, Tanaka GI, Kubuki Y, et al. Role of HTLV-1 proviral DNA load and clonality in the development of adult T-cell leukemia/lymphoma in asymptomatic carriers. Int J Cancer. 2004;110(4):621–5.
14. Manns A, Miley WJ, Wilks RJ, Morgan OSC, Hanchard B, Wharfe G, et al. Quantitative Proviral DNA and Antibody Levels in the Natural History of HTLV-I Infection. J Infect Dis. 1999;180(5):1487–93.
15. Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, Mesnard J-M. The Complementary Strand of the Human T-Cell Leukemia Virus Type 1 RNA Genome Encodes a bZIP Transcription Factor That Down-Regulates Viral Transcription. J Virol. 2002;76(24):12813–22.
16. Mitobe Y, Yasunaga JI, Furuta R, Matsuoka M. HTLV-1 bZIP Factor RNA and protein impart distinct functions on t-cell proliferation and survival. Cancer Res. 2015;75(19):4143–52.
17. Rocha C, Barreto F. Avaliação Da Proteína Hbz No Desenvolvimento Da

Leucemia/Linfoma De Células T Do Adulto: Uma Revisão Sistemática. Saúde.com. 2018;14(4):1303–9.

18. Gallo RC. Research and discovery of the first human cancer virus, HTLV-1. Best Pract Res Clin Haematol [Internet]. 2011;24(4):559–65. Available from: http://dx.doi.org/10.1016/j.beha.2011.09.012

19. Moafi M, Rezvan H, Sherkat R, Taleban R, Asilian A, Esfahani SHZ, et al. Evaluation of IL-12RB1, IL-12B, CXCR-3 and IL-17a expression in cases affected by a non-healing form of cutaneous leishmaniasis: An observational study design. BMJ Open. 2017;7(1):1–9.

20. Yoshida M, Satou Y, Yasunaga J, Fujisawa J, Matsuoka M. Transcriptional Control of Spliced and Unspliced Human T-Cell Leukemia Virus Type 1 bZIP Factor (HBZ) Gene. J Virol. 2008;82(19):9359–68.

21. Murata K, Hayashibara T, Sugahara K, Uemura A, Yamaguchi T, Harasawa H, et al. A Novel Alternative Splicing Isoform of Human T-Cell Leukemia Virus Type 1 bZIP Factor (HBZ-SI) Targets Distinct Subnuclear Localization. J Virol. 2006;80(5):2495–505.

22. Baratella M, Forlani G, Accolla RS. HTLV-1 HBZ viral protein: A key player in HTLV-1 mediated diseases. Front Microbiol. 2017;8(DEC):1–6.

23. Baratella M, Forlani G, Raval GU, Tedeschi A, Gout O, Gessain A, et al. Cytoplasmic Localization of HTLV-1 HBZ Protein: A Biomarker of HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP). PLoS Negl Trop Dis. 2017;11(1):1–15.

24. Borba MMN, Farre L, Bittencourt AL, De Castro-Amarante MF, Galvão-Castro B, Santos LA, et al. Assessment of Genetic Diversity of HTLV-1 ORF-I Sequences Collected from Patients with Different Clinical Profiles. AIDS Res Hum Retroviruses. 2019;35(9):881–4.

25. Mota-Miranda ACA, Barreto FK, Baptista E, Farre-Vale L, Monteiro-Cunha JP, Galvão-

Castro B, et al. Molecular study of HBZ and gp21 human t cell leukemia virus type 1 proteins isolated from different clinical profile infected individuals. AIDS Res Hum Retroviruses. 2013;29(10):1370–2.

26. Clerc I, Polakowski N, André-Arpin C, Cook P, Barbeau B, Mesnard JM, et al. An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ. J Biol Chem. 2008;283(35):23903–13.

27. Tanaka-Nakanishi A, Yasunaga JI, Takai K, Matsuoka M. HTLV-1 bZIP factor suppresses apoptosis by attenuating the function of FoxO3a and altering its localization. Cancer Res. 2014;74(1):188–200.

28. O'Connor L, Strasser A, O'Reilly LA, Hausmann G, Adams JM, Cory S, et al. Bim: A novel member of the Bcl-2 family that promotes apoptosis. EMBO J. 1998;17(2):384–95.

29. Dissinger N, Shkriabai N, Hess S, Al-Saleem J, Kvaratskhelia M, Green PL. Identification and characterization of HTLV-1 HBZ post-translational modifications. PLoS One. 2014;9(11):1–10.

30. Mota ACA, Van Dooren S, Fernandes FMDC, Pereira SA, Queiroz ATL, Gallazzi VO, et al. The close relationship between South African and Latin American HTLV type 1 strains corroborated in a molecular epidemiological study of the HTLV type 1 isolates from a blood donor cohort. AIDS Res Hum Retroviruses. 2007;23(4):503–7.

31. Rego FFDA, Alcantara LCJ, Neto JPM, Miranda ACAM, Pereira ODS, Gonçalves MDS, et al. HTLV type 1 molecular study in Brazilian villages with African characteristics giving support to the post-Columbian introduction hypothesis. AIDS Res Hum Retroviruses. 2008;24(5):673–7.

32. Forlani G, Baratella M, Tedeschi A, Pique C, Jacobson S, Accolla RS. HTLV-1 HBZ protein resides exclusively in the cytoplasm of infected cells in asymptomatic carriers and HAM/TSP patients. Front Microbiol. 2019;10(APR).

Corresponding author

Name: Luciane Amorim Santos

Affiliation: Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia, Brasil; Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brasil.

Email address: lucianeamorim@bahiana.edu.br

Figure 1. Physicochemical profile analysis of R121Q mutation, in red, versus wild-type, in black

Figure 2. Physicochemical profile analysis of V17M mutation, in red, versus wild-type, in black.

Figure 3. Rooted Maximum Likelihood tree of 14 HTLV-1 sequences based on a 760pb LTR region. ML tree with 14 new sequences (names are in bold) and 36 references based on a 760pb LTR region. Sequence L02534 (Isolated Mel 5) was used as an outgroup to root the tree.

Table 1. Common hbz mutations and their frequency between CHTLV and database datasets for each clinical group.

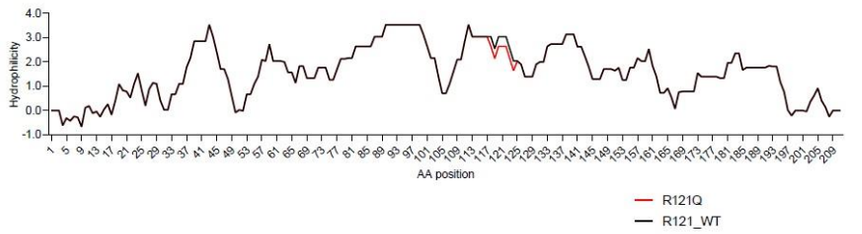
| Nucleotide Position* | Nucleotide Change | AA change | Variant Frequency CHTLV | | Variant Frequency Database | | | |
|---|---|---|---|---|---|---|---|---|
| | | | HAM/ TSP (n=13) | AC (n=2) | HAM/ TSP (n=284) | AC (n=209) | ATLL (n=67) | IDH (n=11) |
| 613 | G → A | A9V | 100.0% | 100.0% | 97.8% | 100.0% | 90.0% | 100.0% |
| 596 | A → G | ** | 84.6% | 100.0% | 28.6% | 65.4% | 30.0% | 80.0% |
| 587 | A → G | S18P | 100.0% | 100.0% | 98.0% | 100.0% | 90.0% | 100.0% |
| 501 | T → C | ** | 84.6% | 100.0% | 7.5% | 20.8% | 9.8% | 80.0% |
| 477 | G → A | ** | *** | 50.0% | 1.9% | 2.8% | *** | *** |
| 447 | T → C | ** | 76.9% | 100.0% | 18.9% | 45.4% | 17.9% | 90.9% |
| 363 | T → C | ** | 76.9% | 100.0% | 19.3% | 44.5% | 17.9% | 81.8% |
| 357 | T → C | ** | 15.4% | *** | 5.3% | 9.6% | 1.5% | *** |
| 327-328 | GG → AA | T104I | 100.0% | 100.0% | 100.0% | 97.1% | 98.5% | 100.0% |
| 312 | T → C | ** | *** | 50.0% | 1.1% | 0.5% | *** | *** |
| 277 | C → T | R121Q | 53.8% | 50.0% | 14.4% | 29.2% | 13.4% | 18.2% |
| 261 | G → A | ** | 23.0% | *** | 6.7% | 12.4% | 1.5% | *** |
| 249 | A → G | ** | 100.0% | 100.0% | 98.6% | 96.2% | 98.5% | *** |
| 248 | T → C | R131G | 7.7% | *** | *** | 0.5% | *** | *** |
| 248 | T → G | ** | 92.3% | 100.0% | 55.4% | 74.2% | 41.8% | *** |
| 243 | C → T | ** | 100.0% | 100.0% | 53.7% | 71.8% | 41.8% | 100.0% |
| 147 | C → T | ** | 84.6% | 100.0% | 11.5% | 31.1% | 11.5% | 80.0% |
| 86 | T → G | ** | 76.9% | 100.0% | 24.4% | 50.0% | 30.0% | 80.0% |
| 57 | C → T | ** | 100.0% | 100.0% | 95.6% | 90.0% | 65.0% | 100.0% |

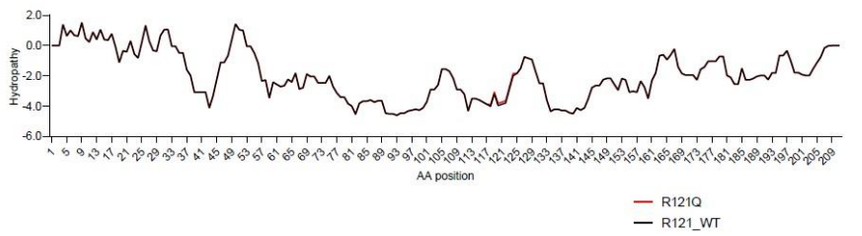
*Position based on hbz region from ATK1 (J02029.1).
**Synonymous mutations with no amino acid change.
***Mutation not found in this clinical status.
AA: amino acid. AC: asymptomatic carrier. ATL: Adult T-cell Leukemia. CHTLV: Centro de Atendimento ao Portador do HTLV. HAM/TSP: HTLV-1-associated myelopathy/tropical spastic paraparesis. IDH: Infective Dermatitis Associated with HTLV-1.

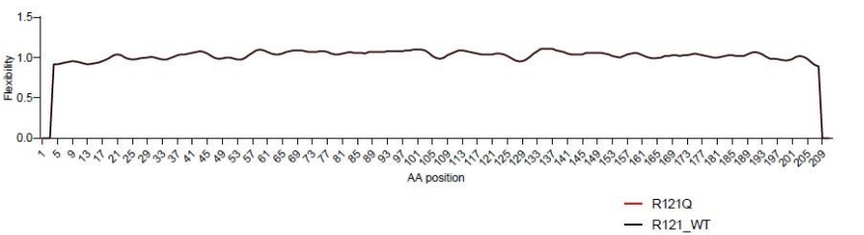
Table 2. Frequency of transcription factors binding sites modification due to mutations in six CTHLV LTR sequences.

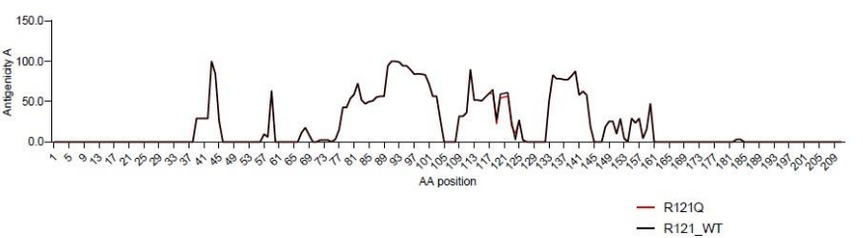
| Mutation position | *Change* | *TFBS by TFSCAN* | *TFBS change* | *CHTLV (n=6)* |
|---|---|---|---|---|
| 178 | G → A | TFIID (173-178) | Creation | 1/5 |
| 275 | T → C | ATF (270-275) | Abrogation | 1/6 |
| 331 | C → T | gammaCAC2 (227-331) | Abrogation | 1/6 |
| 331 | C → T | CACCC-binding factor (327-331) | Abrogation | 1/6 |
| 331 | C → T | CACCC-binding factor (331-335) | Abrogation | 1/6 |
| 418 | G → A | Sp1 (418-424) | Abrogation | 1/6 |
| 418 | G → A | ER (418-424) | Creation | 1/6 |
| 475 | C → T | CACCC-binding factor (472-476) | Creation | 1/6 |
| 477 | C → T | C-Est-2 (477-482) | Creation | 1/6 |

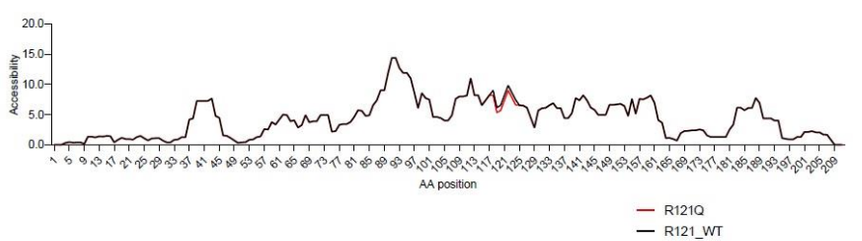
TFBS: transcription factors binding sites.

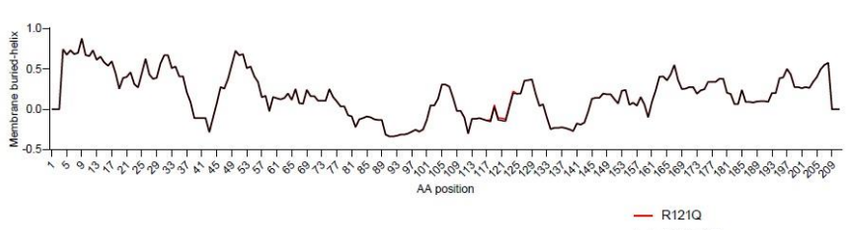

Hydrophilicity
AA position
R121Q
R121_WT

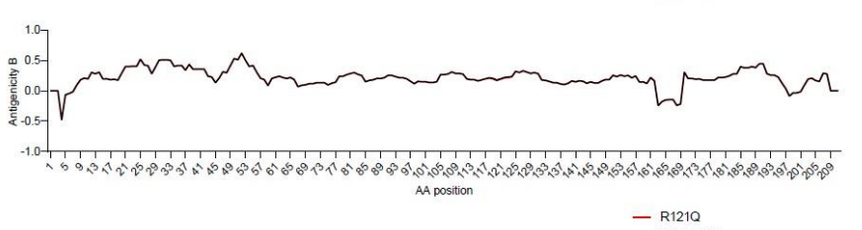

Hydropathy
AA position
R121Q
R121_WT

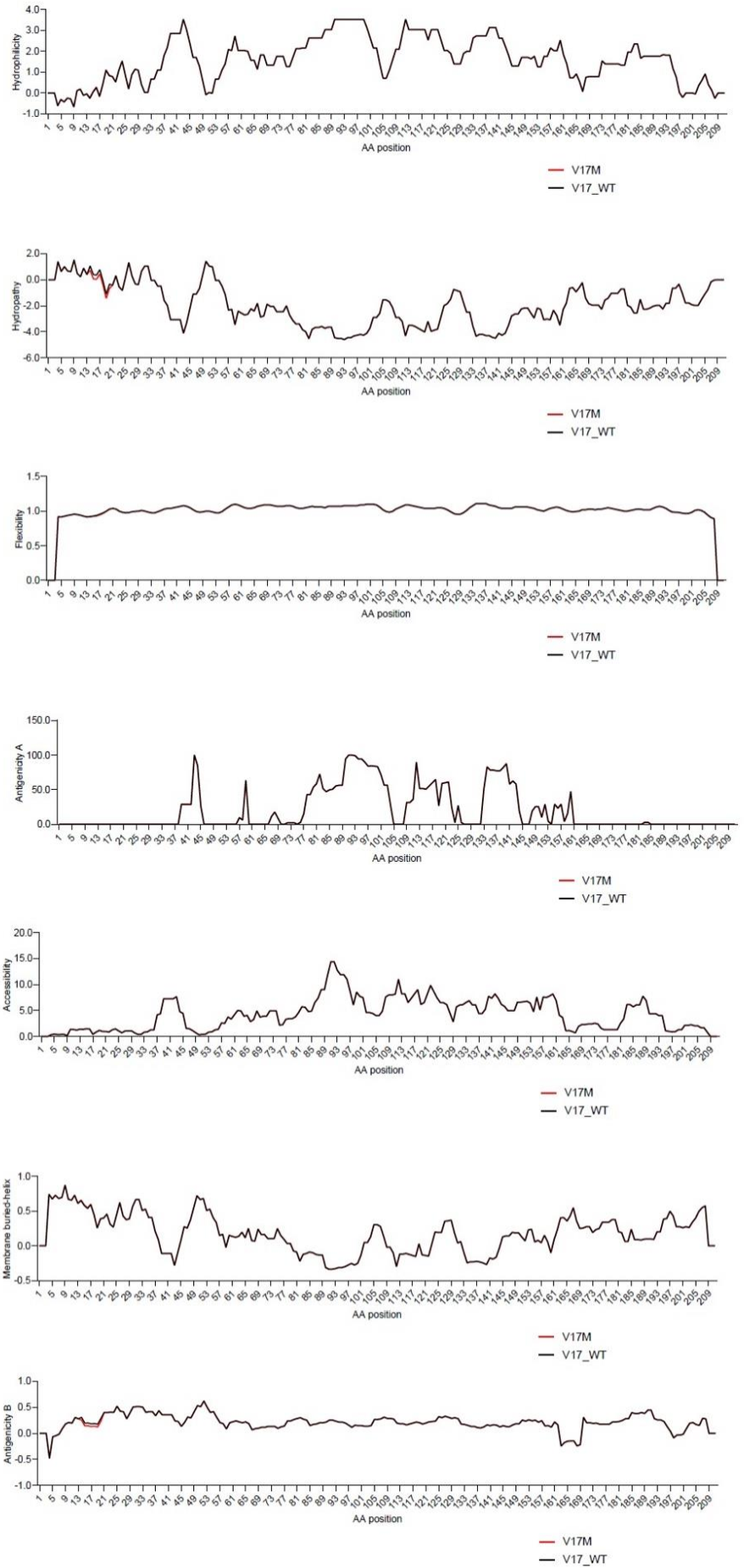

Flexibility
AA position
R121Q
R121_WT

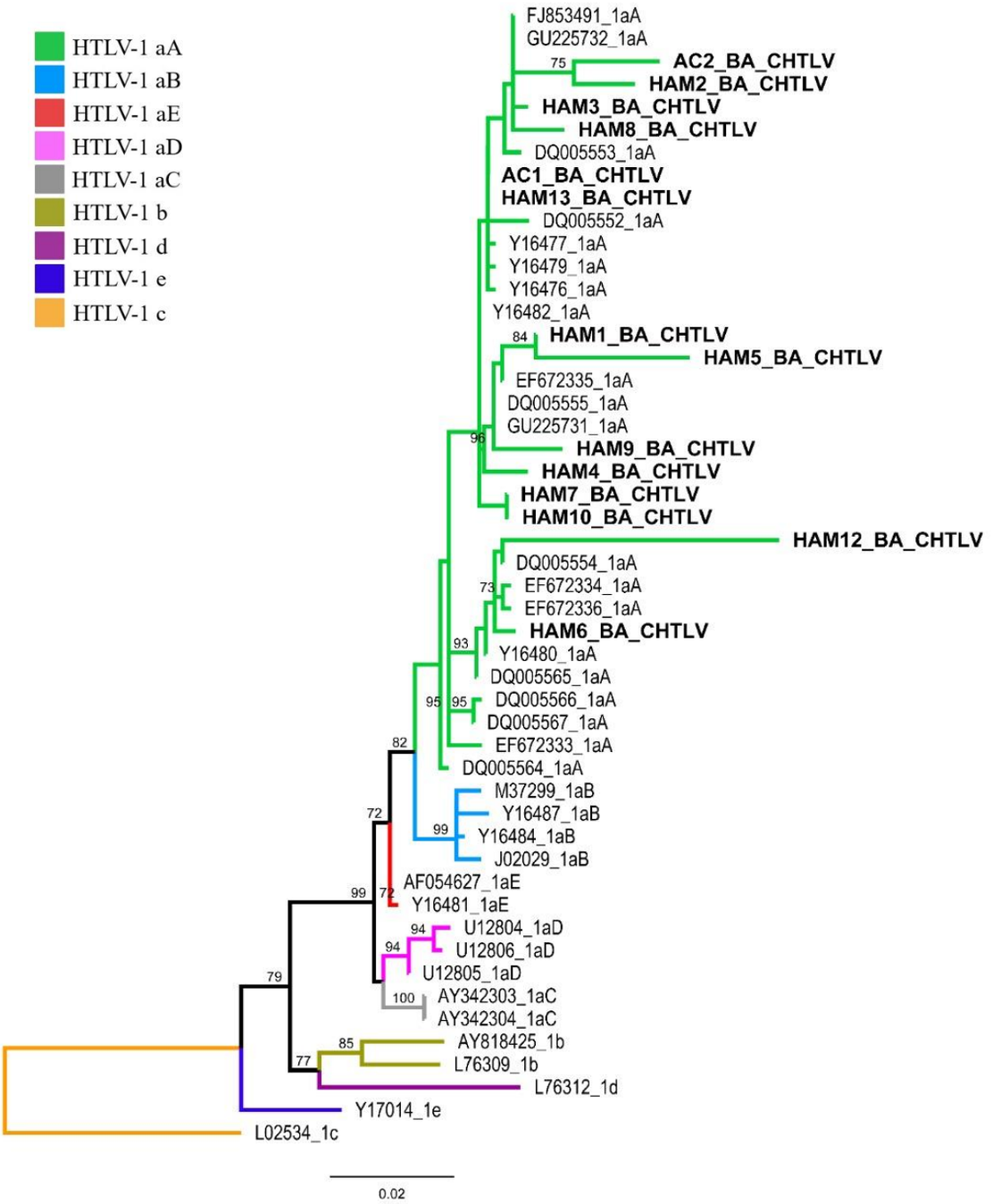

Antigenicity A
AA position
R121Q
R121_WT


Accessibility
AA position
R121Q
R121_WT


Membrane buried-helix
AA position
R121Q
R121_WT


Antigenicity B
AA position
R121Q
R121_WT


Hydrophilicity
AA position
V17M
V17_WT
Hydropathy
AA position
V17M
V17_WT
Flexibility
AA position
V17M
V17_WT
Antigenicity A
AA position
V17M
V17_WT
Accessibility
AA position
V17M
V17_WT
Membrane buried-helix
AA position
V17M
V17_WT
Antigenicity B
AA position
V17M
V17_WT


HTLV-1 aA
HTLV-1 aB
HTLV-1 aE
HTLV-1 aD
HTLV-1 aC
HTLV-1 b
HTLV-1 d
HTLV-1 e
HTLV-1 c
FJ853491_1aA
GU225732_1aA
AC2_BA_CHTLV
HAM2_BA_CHTLV
HAM3_BA_CHTLV
HAM8_BA_CHTLV
DQ005553_1aA
AC1_BA_CHTLV
HAM13_BA_CHTLV
DQ005552_1aA
Y16477_1aA
Y16479_1aA
Y16476_1aA
Y16482_1aA
HAM1_BA_CHTLV
HAM5_BA_CHTLV
EF672335_1aA
DQ005555_1aA
GU225731_1aA
HAM9_BA_CHTLV
HAM4_BA_CHTLV
HAM7_BA_CHTLV
HAM10_BA_CHTLV
HAM12_BA_CHTLV
DQ005554_1aA
EF672334_1aA
EF672336_1aA
HAM6_BA_CHTLV
Y16480_1aA
DQ005565_1aA
DQ005566_1aA
DQ005567_1aA
EF672333_1aA
DQ005564_1aA
M37299_1aB
Y16487_1aB
Y16484_1aB
J02029_1aB
AF054627_1aE
Y16481_1aE
U12804_1aD
U12806_1aD
U12805_1aD
AY342303_1aC
AY342304_1aC
AY818425_1b
L76309_1b
L76312_1d
Y17014_1e
L02534_1c
75
84
96
73
93
95
95
82
72
99
99
72
94
94
79
100
85
77
0.02

## 2. Proposta de Submissão

**2.1.** Revista: AIDS Research and Human Retroviruses

**2.2.** Instruções:

**2.2.1.** File Naming: All file names should be in English and contain only alphanumeric characters. Do not include spaces, symbols, special characters, dashes, dots, or underscores. Label each file with the first author's last name, followed by the content of the file (ie: SmithText; SmithTables; SmithFigureLegends; SmithFig1, etc). ***Important:*** Please upload individual files of all manuscript material as described herein — please do NOT upload a combined PDF file containing all material in your submission.

**2.2.2.** Preparation of Manuscript: Prepare text of manuscripts, figure legends, and tables in Microsoft Word, double spaced. The order of elements in each manuscript should be:

**2.2.2.1.** Title page (with full manuscript title, all contributing authors and their affiliations, a short running title, a denotation of the corresponding author, and a list of 3-6 keywords

**2.2.2.2.** Abstract

**2.2.2.3.** Main text (do not embed figures or tables)

**2.2.2.4.** Conclusion (as a separate paragraph, not as part of the Discussion section)

**2.2.2.5.** Acknowledgments (if applicable)

**2.2.2.6.** Sequence Data / Accession Number(s)

**2.2.2.7.** Authorship confirmation statement

**2.2.2.8.** Author(s') disclosure statement(s)

**2.2.2.9.** Funding statement (see below)

**2.2.2.10.** References

**2.2.2.11.** Figure legends

**2.2.2.12.** Tables

**2.2.2.13.** Supplemental files (if applicable; supplied in Word, not as a PDF)

**2.2.3.** Manuscript Text: Maximum word count should not exceed 4,000 words. In general, the text should be organized under the headings: Introduction, Materials and Methods, Results, Discussion, and Conclusions. Use only standard abbreviations, which can be found in the AMA's Manual for Authors & Editors or the Council of Biology Editors Style Manual, 10/e. At first usage, spell out terms and provide abbreviations in parentheses. Thereafter, use only the abbreviations. It is not necessary to spell out standard units of measure. Use generic names for drugs if possible. If you wish to use a proprietary drug name the first time it appears, use the generic name followed by the proprietary name, manufacturer, and location in parentheses.

**2.2.4.** References: References must be prepared in word, double spaced, and numbered consecutively as they are cited in the text (using superscript numbers). Include reference section as part of the main text file, not as a separate file. References appearing for the first time in tables and figures must be numbered in sequence with those cited in the text where the table or figure is mentioned. Use journal abbreviations as provided by PubMed/Medline. List all authors when there are six or less. When there are more than six, list the first three, followed by et al. If references to personal communications or unpublished data are used, they are not to be in the list of references. They should be referred to in the text in parentheses: (AB Jones, personal communication). Include among the references any papers that have been accepted but have not yet published; identify the name of publication and add "In Press." If the reference has been published online, provide the DOI number in place of the page range. Sample style for references:

**2.2.4.1.** *Journal article:* Hendrix CW,Andrade A, Bumpus NN, Kashuba AD, Marzinke MA, Moore A et al. Dose frequency ranging pharmacokinetic study

of tenofovir-emtricitabine after directly observed dosing in healthy volunteers to establish adherence benchmarks (HPTN 066). AIDS Res Hum Retroviruses 2015;32(1):32-43.

**2.2.4.2.** *Book:* Vicenzi E, Poli G. Human Retroviruses: Methods and Protocols (Methods in Molecular Biology). Totowa, NJ: Human Press (Springer) 2014.

**2.2.4.3.** *Chapter in a book:* Cimarelli A, Darlix J-L. HIV-1 Reverse transcription. In: Human Retroviruses: Methods and Protocols (Methods in Molecular Biology), (Vicenzi E, Poli G.,eds.) Totowa, NJ: Human Press (Springer) 2014; pp. 55-70.

**2.2.5.** Legends: Figure legends should be uploaded as a separate Word file and double spaced. In the legend, provide explanations for any abbreviations, arrows, etc. that appear in the figure. If the illustration is taken from a copyrighted publication, permission must be secured, appropriate credit must be given in the legend, and a corresponding reference must appear in the reference section.

**2.2.6.** Tables: All tables should be prepared in one single Word file. Provide a title for each table. Cite tables in sequence in the text. Explain abbreviations used in the body of the table in footnotes. If the table is taken from a copyrighted publication, permission must be secured, appropriate credit must be given in the legend, and a corresponding reference must appear in the reference section.

**2.2.7.** Figures: Submission of high resolution .TIFF or .EPS figure files is strongly recommended.

**2.2.7.1.** Figures should not be embedded within the manuscript file.

**2.2.7.2.** Cite figures consecutively in text within parentheses.

**2.2.7.3.** A legend should be supplied for each figure and all legends numbered consecutively.

**2.2.7.4.** Images should not show the name of a patient or a manufacturer.

**2.2.7.5.** Do not include any illustrations as part of your text file.

**2.2.8.** Correspondence Address: Following the references, provide the name and complete affiliation and e-mail address of the person to whom correspondence should be sent.