

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341908419>

Evolutionary Temporal Footprint of an HIV-1 Envelope Protein in an Epidemiologically Linked Cluster

Article in The Open AIDS Journal · May 2020

DOI: 10.2174/1874613602014010041

CITATIONS

0

READS

40

5 authors, including:



Élidamar Nunes

University of São Paulo

8 PUBLICATIONS 35 CITATIONS

[SEE PROFILE](#)



Muhammad Shoaib Arif

Northwestern University Chicago

22 PUBLICATIONS 121 CITATIONS

[SEE PROFILE](#)



J.R.C. Piqueira

University of São Paulo

249 PUBLICATIONS 2,095 CITATIONS

[SEE PROFILE](#)



The Open AIDS Journal

Content list available at: <https://openaidsjournal.com>



RESEARCH ARTICLE

Evolutionary Temporal Footprint of an HIV-1 Envelope Protein in an Epidemiologically Linked Cluster

Elidamar Nunes de Carvalho Lima^{1,2,*}, Rodrigo Sucupira Andrade Lima¹, Muhammad Shoaib Arif², José Roberto Castilho Piqueira¹ and Ricardo Sobhie Diaz²

¹Department of Telecommunication and Control Engineering, Engineering School, University of Sao Paulo, Sao Paulo, Brazil

²Department of Medicine, Infectious Diseases Division, Federal University of Sao Paulo, Sao Paulo, Brazil

Abstract:

Background:

The C2V3C3 region of gp 120, encoded by the HIV-1 envelope gene (*env*), is an important antigenic target, a key determinant for viral evolution and essential for determining epitopes for vaccines.

Methods:

The relationships among genetic sequence diversity, selective pressure, constraints on HIV-1 envelope protein were explored and also correlated this analysis with information entropy; hypermutation; HIV tropism; CD4⁺ T cell counts or HIV viral load. A total of 179 HIV-1 C2V3C3 sequences derived from cell-free plasma were used, determined from serial samples, in four epidemiologically linked individuals (one infected blood donor, two transfusion recipients and a sexual partner infected by one of the recipients) over a maximum period of 8 years. This study is important because it considers the analysis of patterns in genomic sequences, without drugs and over time.

Results:

A temporal relationship among information entropy, hypermutation, tropism switch, viral load, and CD4⁺ T cell count was determined. Changes in information entropy were time-dependent, and an increase in entropy was observed in the C2V3C3 region at amino acids G313 and F317-I320 (related to the GPGR-motif and coreceptor tropism), and at amino acids A281 in C2 and A346 in C3, related to immune escape.

Conclusion:

The increase of information entropy over time was correlated with hypermutation and the emergence of nonR5- strains, which are both associated with more variable genomes.

Keywords: Information entropy, HIV-1, C2V3C3 region, Genetic variability, Genome evolution, Immune escape.

Article History

Received: January 08, 2020

Revised: March 30, 2020

Accepted: April 11, 2020

HIGHLIGHTS

✓ Increased information entropy correlates with a tropism switch from R5 to non-R5 strains.

✓ Increased information entropy correlates with decreased CD4⁺T cell count and viral load, and increased hypermutation.

✓ HIV genetic evolution and the selection of escape mutants result in the emergence of more disorganized HIV genomes.

* Address correspondence to this author at the Department of Telecommunication and Control Engineering, Engineering School, University of Sao Paulo, Sao Paulo, Brazil; Tel: +55 (11) 3091-5647, +55 (11) 96326-5550; E-mail: elidamamunes@gmail.com

1. INTRODUCTION

The ability of HIV-1 to rapidly evolve is one of its most striking features, enabling its immune escape and favoring its persistence in an infected host [1, 2]. The HIV-1 envelope gene (*env*) encodes the viral envelope glycoprotein, which is a heavily glycosylated trimer of non-covalently linked, heterodimeric glycoproteins composed of the surface-exposed gp120 and the transmembrane gp41. Extensive N-linked glycosylation of the envelope glycoprotein effectively shields many of its conserved epitopes from antibody recognition. Based on sequence variability, gp120 is divided into five conserved regions (C1-C5) and five hypervariable loops (V1-V5). Hypervariable loop 3 (V3) of gp120 is approximately 35 amino

acids long, frequently glycosylated, and highly variable and has a disulfide-bonded structure that has potential impact on several functions of the envelope protein. The V3 loop harbors the most important determinants of viral tropism (*i.e.*, coreceptor usage), as well as major antigenic neutralizing epitopes [3 - 6].

The relevance and novelty of this study lie in the analysis of patterns in genomic sequences, by using information metrics as well as the variability accumulated occurrence, in the natural history of HIV infection over time, in an epidemiologically linked cluster once a few studies address this kind of approach. Thus, the analysis using informational metrics lies in quantifying the genomic patterns over time, which is fundamental to study and planning of new targets based on epitopes behavior. In this sense, the present study analyzed the inpatient evolution of the gp120 C2V3C3 region in four distinct epidemiologically linked individuals over the course of approximately 7 years, focusing on evolutionary selective pressure and its association with information entropy, hypermutation, tropism and clinical markers of disease progression [7 - 11]. When used in genomic analysis, information entropy brings important insights into the evolutionary relationship between the virus and its host, which is essential for designing new antiretroviral drugs and vaccine candidates against HIV, and also provides insights into virus persistence. Since genomic evolution is a process that enables escape from environmental constraints imposed by the immune system or antiretroviral [12 - 18], we evaluated information entropy over time and its association with HIV-1 evolution as inferred from genetic diversity.

2. METHODS

2.1. Patients and Sequences

The sequences used in this study were obtained from a cluster of four HIV-1-infected individuals comprising a blood donor, two recipients of infected blood from the donor, and a sexual partner infected by one of the recipients [19]. The patients enrolled in this study were at an asymptomatic stage of infection and were antiretroviral-naïve adults. Sequence data were obtained from serum from the blood donation sample and from plasma samples collected from the four members of the epidemiologic cluster over time. Sequences derived from cell-free plasma were determined from serially collected samples. Briefly, viral RNA from plasma samples collected from the four members of the epidemiological cluster over time was used to generate molecular clones by single genome PCR amplification from viruses isolated at distinct time-points, and the sequences were generated from a section of *env* that includes the 3' end of the C2-encoding region, the V3-encoding region, and the 5' end of the region that encodes C3. Nine or more sequences per patient per time-point were obtained, which yielded a total of 179 sequences. The number of samples was not constant; a range of 7-11 clones per segment were sequenced to obtain a total of approximately 49 sequences per patient. The C2V3C3 *env* region HIV-1 subtype B sequence data were from samples collected over the 8-year period from 1986 to 1993, and the analyzed *env* fragment encodes 76 residues corresponding to amino acids F277 to F353 of the

HXB2 reference strain. The sequences are from an epidemiologic cluster that is composed of a blood donor (D.O.) who donated HIV-infected blood in 1985, two recipients (R.A. and R.B.) infected by the contaminated blood, and one individual who was infected by sexual contact (S.C.) with R.B. These individuals remained antiretroviral naïve during the entire follow up of seven years; thus, the molecular evolution depicts natural progression in the host. The individuals remained asymptomatic during the course of the study and had stable and normal CD4⁺ and CD8⁺ counts and low viral loads. HIV-1 plasma viremia was quantified with a quantitative competitive RT-PCR assay as we can see on [19]. All sequences reported in this publication were submitted to GenBank (accession numbers U29433–U29437, U29956, U29957, U29959–U30074, U30077–U30145, U31573–U31582, and U43035–U43054). For further analysis, the sequences of each patient were grouped according to collection year and aligned using *ClustalW* [20] with sequence D85.40 (from D.O.) as a reference because it represents the strains exclusively found in the blood inoculum. Sequencing errors, genomic regions with deletions or insertions were excluded from the analysis to preserve the reading frame.

2.2. Evolutionary Analysis

We analyzed the evolution pattern of the C2V3C3 regions of HIV-1 gp120 over time by determining information entropy and its relationship with tropism, hypermutation, viral load, and CD4⁺ T cell count. Entropy is a standard measure to evaluate protein variability that quantifies the uncertainty of information per site or position in the genome and considers the number of possible amino acid replacements and their frequency. Information entropy (*H*) is commonly used to quantify the uncertainty of information about the amino acid or protein at a given position and their fixation over time. The classical *Shannon* formula for the entropy, or information content, per position of the amino acid sequence, is written as:

Equation 1: Information Entropy [21].

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

where $p(x)$ is the probability of the base (A, T, C, G) in the given sequence; in other words, $p(A)$ is the probability of the occurrence of the base "A". The probability was estimated by the frequency (by counting the occurrence) of each amino acid (a.a.), x represents each of the possible a.a. and \log_2 represents the logarithm in bits. A value of 0 for H indicates that all sequences are identical at a given position, whereas a non-zero value indicates that different amino acids are present. The average information entropy over time at each position in the C2V3C3 regions of HIV-1 gp120 was determined using MATLAB software [22]. An information entropy value less than 0.2 was considered conserved, a value between 0.2-0.5 was considered semi conserved, and a value above 0.5 was considered nonconserved [23, 24]. Hypermutation/G→A substitution analysis of the HIV sequences was performed using the Hypermut 2.0 tool, available at LANL HIV Sequence Database [25], and the coreceptor tropism of the sequences was

evaluated by the Geno2pheno algorithm, available at <http://coreceptor.geno2pheno.org/index.php> [26], with a False Positive Rate (FPR) cutoff of 10% in accordance with European guidelines [27].

2.3. Analysis and Data Validation

To quantify the relationships among the evolutive patterns of the envelope genes of the group, *Pearson's* correlation coefficient was calculated for the time series data. The relationship among information entropy, CD4⁺T cell count, viral load, tropism, and hypermutation in the group was calculated by pairwise correlations matching each genome position, and the corresponding *P* value of each measurement, hypothesis test, entropy means, and R2 were determined. Statistical analysis was performed using Excel®, MATLAB and GraphPad Prism version 6.0 software [22, 28]. We considered a *P* value of 0.05 to be statistically significant, aiming to a more conservative analysis.

3. RESULTS

Information entropy showed a progressive increase in all individuals in this cohort over time. A correlation between information entropy and viral load was observed in all individuals but was statistically significant only in D.O. ($p < 0.007$) (Fig. 1, Panel 1). Similarly, information entropy correlated with CD4⁺T cell count in patients D.O. ($p = 0.223$), R.A. ($p = 0.490$), R.B. ($p = 0.537$), and S.C. ($p = 0.604$). Additionally, R.B. presented a decrease in CD4⁺T cell count associated with an increase in entropy (Fig. 1, Panel 2). To obtain further insight into molecular evolution, we analyzed the relationship between information entropy and hypermutation. In an analysis of G-A hypermutation vs entropy, a significantly positive correlation was seen in D.O. ($p = 0.017$) and R.A. ($p = 0.096$), but no statistically significant correlation was seen in R.B. ($p = 0.800$) or S.C. ($p = 0.350$), as in Fig. 1, Panel 3. Only D.O. and R.B. harbored a small proportion of CXCR4-tropic strains (*i.e.*, non-R5), which were present from the beginning of the infection (Fig. 2). We, therefore, inspected the relationship between information entropy and G-A hypermutation and the presence of non-R5 strains. The progressive increases in entropy and hypermutation in D.O.

and R.B. could have resulted in the emergence of non-R5 strains and, in the case of R.B., disease progression. R.A. did not present any AIDS-defining opportunistic infections or neoplasms during the follow-up period, although the CD4⁺T cell decay clearly demonstrated HIV-1 disease progression, whereas R.B. died in 1991 because of AIDS. Additionally, non-R5 strains were identified in D.O. in 1993 and in R.B. as early as the second evaluated time point in 1987 (Fig. 2). Here, we utilized the *Geno2pheno*[coreceptor] algorithm to determine coreceptor usage among the analyzed genomic sequences. We observed an increase in information entropy over time and emergence or increase in the prevalence of non-R5 strains over time when single genome amplification methodology was used. In addition, when analyzing the average information entropy at each amino acid position in C2V3C3 of gp120 computed over time, we observed a persistent increase in information entropy in all individuals (Figs. 3 and 4). Genomic variations differed substantially among individuals at later time points. Table 1 summarizes the amino acid positions that remained conserved (entropy score < 0.20) or were associated with viral evolution (entropy score > 0.5) over time. D.O. showed higher entropy in the V3 region at a.a. K305, R311, P313, R315, F317 and I320 (related to tropism, immune escape and GPGR-motif), in the C3 region at a.a. A346 (related to immune escape) and in the C2 region at a.a. A281 (related to immune escape) (Fig. 3). R.A. showed more entropy in the C3 region at a.a. G313 and S334-A336 (related to GPGR-motif and immune escape) [29, 30]. R.B. showed higher entropy in the V3 region at a.a. A316 and I320 (immune escape and tropism, respectively), as well as in the C2 region at a.a. D279, A281 (immune escape) and in the C3 region at a.a. A346. S.C. showed higher entropy in the C3 fragment at a.a. A346 (immune escape), in the V3 region at a.a. K305, P313 and I320 (related to tropism and GPGR-motif), and in the C2 region at a.a. A281 (related to immune escape) [31]. Similarly, several residues were identified as common in all individuals of the cohort and remained conserved (information entropy equal to zero or < 0.2): a.a. N280, T283-I285, and L288 in C2, a.a. R298, R304, I309, and Q328-I333 in V3, and a.a. E351 and Q352 in C3 [29 - 32].

Table 1. Amino acid positions demonstrating increased information entropy over time in the C2, V3, and C3 regions of HIV-1 gp120 from members of an epidemiologically linked cluster. Positions with high entropy associated with tropism changes are underlined, those associated with immune escape are in red, and those associated the GPGR-motif are boxed. *Information entropy cut off < 0.20 and > 0.50 showed conserved and variable residues, respectively.

Amino Acids		
ID	Information Entropy $< 0.20^*$	Information Entropy $> 0.50^*$
D.O.	D279, N280, T283-L288, V292, I294, P299, N302-R304, R306, I307, I309, Q310, Q328-I333, K337-T341, E351, Q352.	A281, T290, S291, N300, K305, <u>R308</u> , R311, P313, R315, F317, I320, <u>G324</u> , L342, I345, A346, L349, R350, F353.
R.A.	F277-A281, T283-289, I294-R304, G314-T319, G321-I323, Q328-I333, S347-E351.	K282, T290, S291, E293, K305, <u>R308</u> , I309, G313, I320, G324, S334-A336, N339, L342, K343, A346.
R.B.	F277, T278, I284, I285, L288, N289, R298-K305, I309-P313, R315-S334, K343-I345, R350-Q352.	D279, A281, T283, E293, N295, T297, R306, R308, G314, A316, I320, <u>K322</u> , G324, R335, L342, A346, F353.
S.C.	D279, N280, T283-I285, Q287-N289, R298-R304, I309-G312, G314-A316, N325-C331, K337-T341, K343-I345, S347, E351, Q352.	T278, A281, K282, V286, T290, S291, E293, N295, T297, K305, R308, P313, <u>I320</u> , G324, A336, L342, A346, L349, R350, F353.

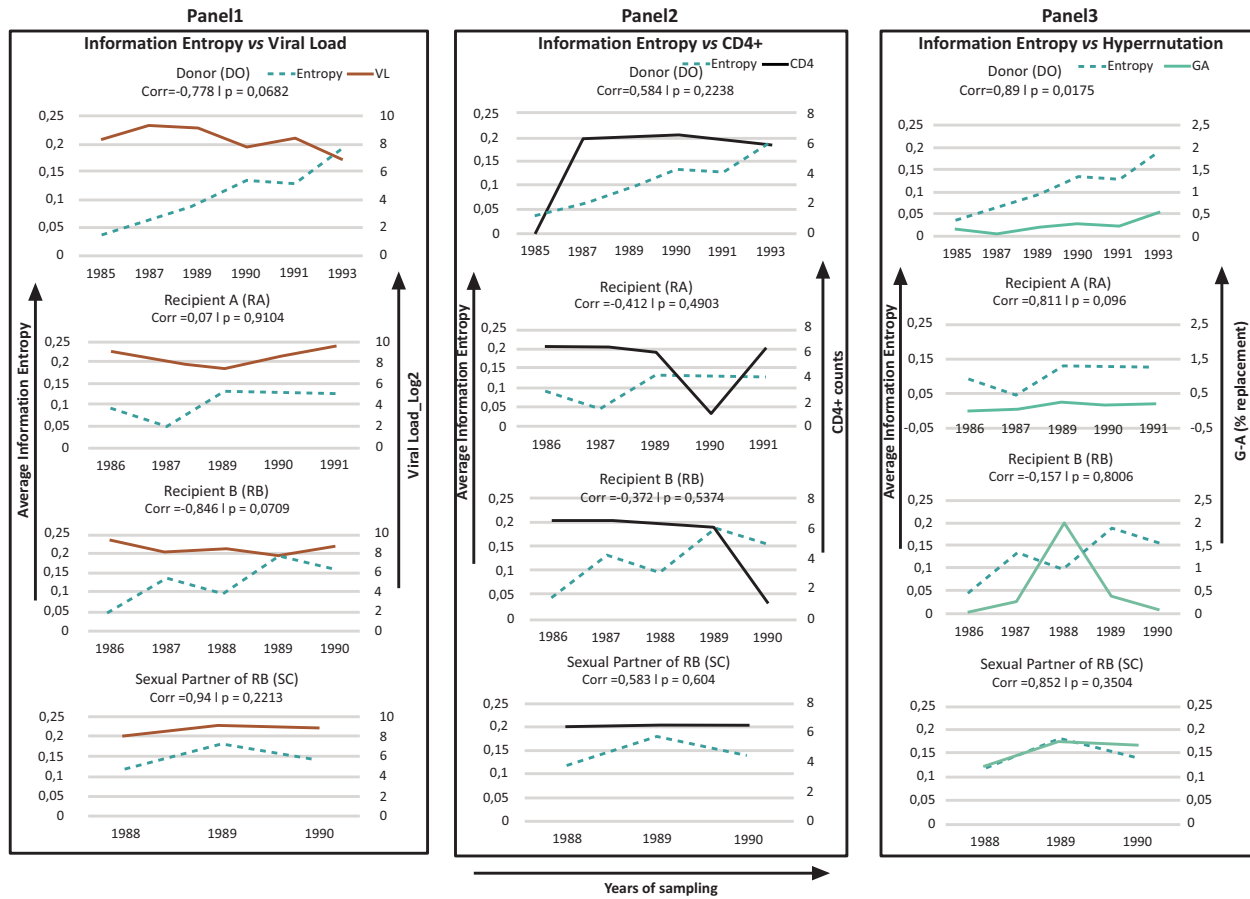


Fig. (1). Correlations of changes in information entropy over time with viral load, CD4+ T cell count, and hypermutation. Variability among sequences grouped according to collection year was calculated as the average information entropy. Correlations of changes in information entropy over time with the plasma viral load (log2 RNA copies/mL) (**Panel 1**), CD4+ T cell count (cells/ μ L) (**Panel 2**), and hypermutation (**Panel 3**) were determined.

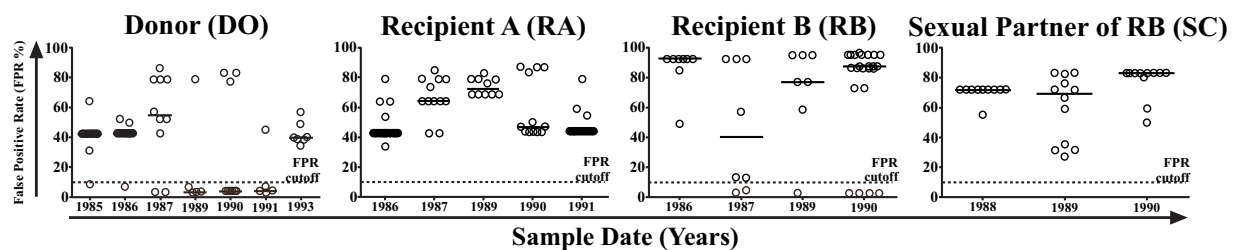


Fig. (2). Determination of coreceptor tropism evolution according to the False Positive Rate (FPR) over time assessed by Geno2pheno [coreceptor] for all individuals in the cluster. Coreceptor tropism was assessed for all strains using the Geno2pheno algorithm with an FPR cutoff set at 10% (indicated as a dotted line in each graph). The FPR is the percentage of all sequences that would be falsely called an X4-capable if one used the FPR at that cutoff. DNA sequences with FPR > 10% are classified as R5 strains (indicated with black open circles), and those with FPR \leq 10% are classified as non-R5 strains (indicated with red open circles).

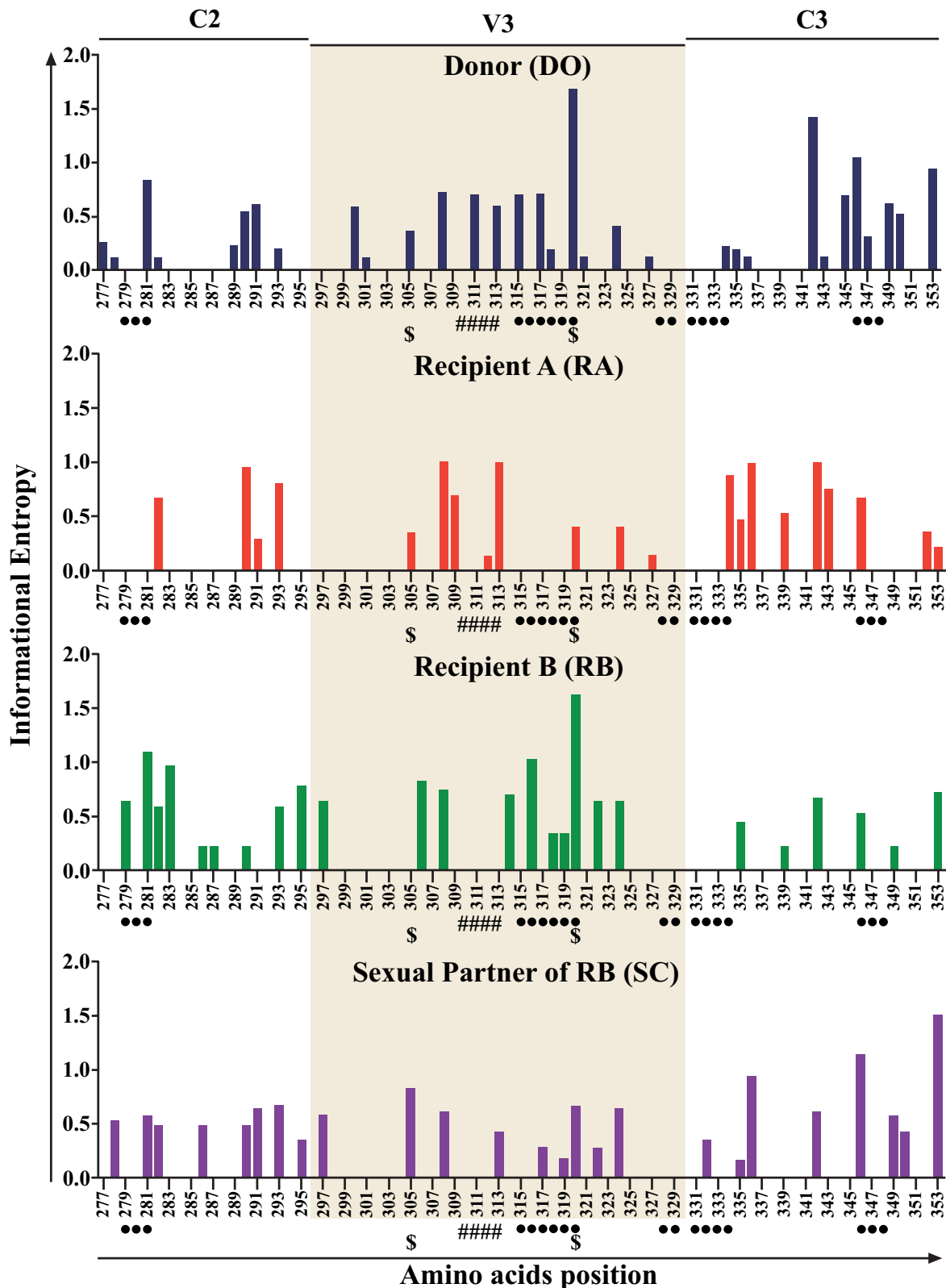


Fig. (3). Average information entropy at each amino acid in C2, V3, and C3 of HIV-1 gp120 for each individual in the epidemiological cohort. The average information entropy at each residue was calculated for the aligned sequences representing seven years of follow-up for each individual in the cohort. Amino acids are numbered according to the gp120 residues of the HIV-1 HXB2 reference strain. The V3 loop is shown as the shaded area, • represents amino acids related to immune escape, # indicates the GPGR motif at the tip of the V3 loop, and \$ indicates the amino acids at residues 11 and 25 of the V3 loop, which associated with the determination of coreceptor tropism.

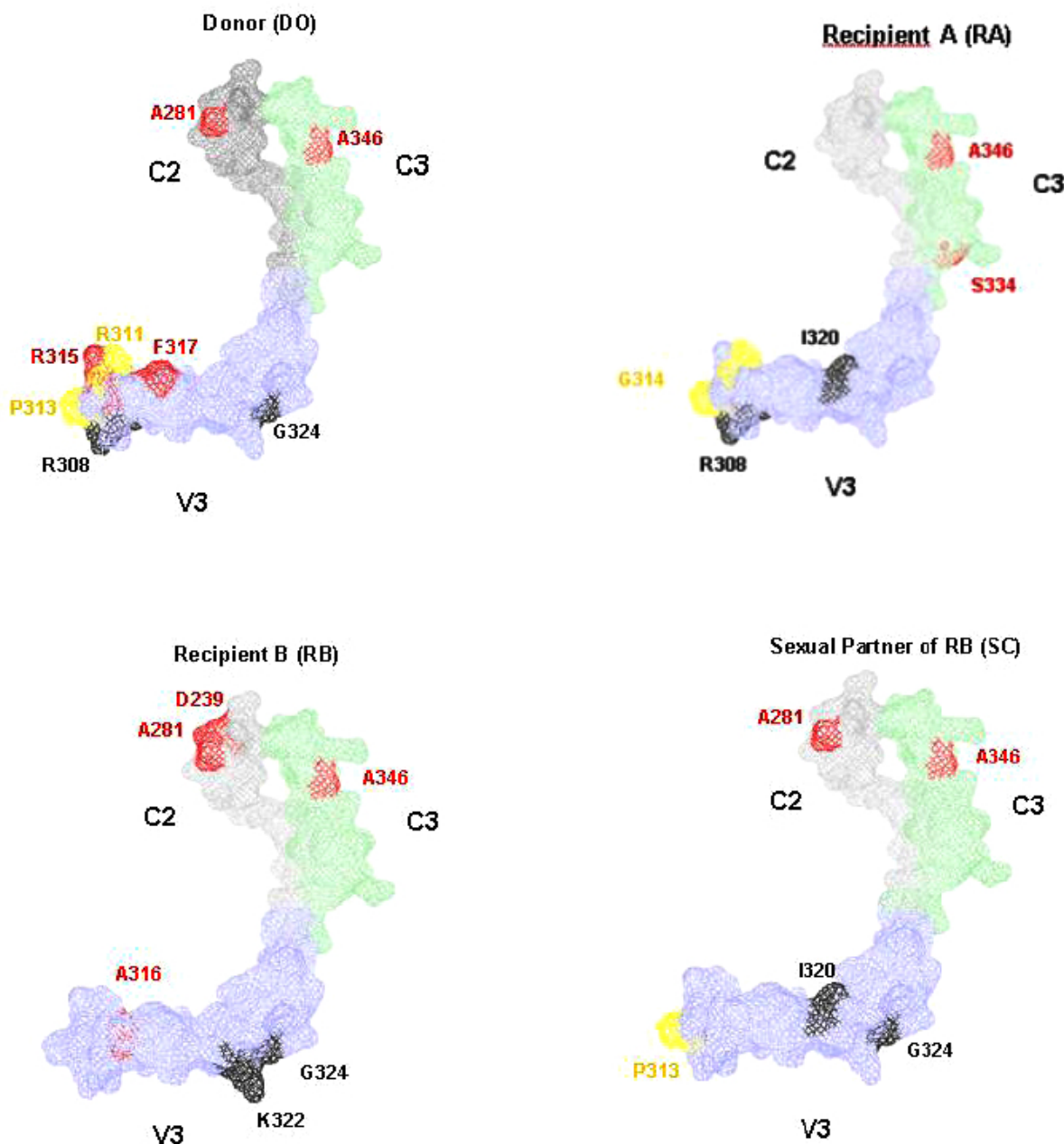


Fig. (4). Conformational structure of C2, V3 and C3 envelope regions in HIV-1 for the epidemiological cohort showing the high entropy variations points on a protein 3D view. The information entropy was calculated at each amino acid in C2 (gray), V3 (blue), and C3 (green) of HIV-1 gp120 for each individual in the epidemiological cohort. Amino acids are numbered according to the gp120 residues of the HIV-1 HXB2 reference strain. For this structural drawing, Swiss PDB viewer v.3.7 software was used, and the X-ray coordinates of the Protein Data Bank PDB are available from <https://swissmodel.expasy.org/> under the access codes 5fyk.1.D. Amino acids high entropy related to immune scape are show in red (A281, D239, A346, R315, A316, A346, F317, S334), and related to tropism are colored in black (R308, I320, K322, G324, I320), aminoacids correspond to GPGR-motif are colored in yellow (R311, P313, G314).

4. DISCUSSION

This study describes the relationship between informational entropy in C2V3C3 of Gp120 and correlates it with APOBEC-driven G-A substitutions, viral load, CD4+T cell and predicted tropism. The informational entropy correlation with CD4+T cell count (Fig. 1, Panel 2) similar to another study, that demonstrated that less organized HIV genomes, as inferred from higher levels of information entropy, correlate with less

competent host immune systems [24]. Non-R5 strains have been associated with the presence of positively charged amino acids in the V3 loop of gp120. It has also been described that, in contrast to CCR5, the CXCR4 coreceptor is filled with negatively charged amino acids, which better enables the attachment of dual-tropic HIV strains [33, 34].

Interestingly, positively charged amino acids, such as arginine and lysine, are usually coded by nucleotide triplets

such as AGA and AGG (arginine) or AAA and AAG (lysine), which have large amounts of As. Non-R5 HIV-1 strains may emerge early in the HIV infection and be associated with faster HIV-1 disease progression [35 - 37]. We utilized the Geno2pheno [coreceptor] algorithm to determine coreceptor usage among the analyzed genomic sequences. Although phenotypic assays are the gold standard for determining coreceptor tropism, genotypic assays provide a reasonable alternative and have been increasingly used. Additionally, the European Consensus Group considers genotypic assays sufficient prior to prescribing maraviroc. Among the different coreceptor tropism prediction algorithms available online, Geno2pheno [coreceptor] has demonstrated comparable performance to the original and enhanced-sensitivity Trofile phenotypic assays [26, 27, 38 - 40].

Additional evidence for the validity of Geno2pheno [coreceptor] predictions is based on the fact that, in an individual patient, the FPR values tend to progressively decay, leading to the emergence of CXCR4-tropic strains with a mean evolution time of 27.29 months (range, 8.90 to 64.62) when population sequencing is used. In addition, increases in entropy over time observed here are indicative of viral evolution with discrete escape and stabilization of the viral population. Amino acids with high entropy scores are located within the core of the envelope protein. Gp120 variations are not uniform; occur at different frequencies and in different fragments of HIV-1 gp120, and allegedly influence protein folding. Conserved residues, as seen in C2V3C3, are responsible for viral structural maintenance and replication (Fig. 3). Interestingly, in all individuals of this cohort, amino acids A281, 282, 290-293, 295, 305, 308, 313, 320, 324, 342, 343, 346, 350, and 353 (related to immune escape, T cell recognition epitopes, and tropism, were constantly under environmental disturbance or in an adaptive process [29 - 32, 40 - 43]. Additionally, with the exception of the result on R.B. (Fig. 1), the increase in entropy is proportional to increases in hypermutation and tropism change; it is also associated with the structural genomic order, and consequently, the structural and/or functional stability of DNA.

Although we have performed analysis using molecular clones in a temporal evolutive fashion, we recognize that this work has a serious limitation in scope since the data set is derived from only 4 HIV infected patients. Nonetheless, based on the obtained results, we can infer that these conserved residues are of interest for future studies aimed at the design of epitopes and the development of therapeutic molecules to treat HIV. In this context, these results highlight the constraints on the evolution of the same viral quasispecies present in distinct human hosts over time.

CONCLUSION

We identified important functional constraints related to evolution in the genomic regions C2V3C3 of HIV-1 gp120. The genomic variability occurred in specific residues related to immune escape was time-dependent among all individuals and showed a direct relationship with viral evolution, escape fitness and disease progression (D279, A281, R315, A316, F317, S334, A346; R308, I320, K322, G324; R311, P313).

Moreover, according to the results seen in one of the patients and considering the literature [36], we can speculate that HIV evolution, leading to the emergence of less organized genomes, will lead to the emergence of hypermutated HIV strains and, consequently, more cytopathic non-R5 strains develop, potentially contributing to a rapid disease progression for HIV.

Thus, understanding the evolutionary pattern of HIV-1 is critical, and genomic analysis using information entropy provides important insights into evolution and genetic restriction and the conserved regions identified in this study that can inspire designers of new molecules and vaccines based on conservation patterns.

AUTHORS' CONTRIBUTIONS

Designed the research: E.N.d.C.L., R.S.D., and J.R.C.P. Performed and interpreted the experiments in evolutionary analysis and information entropy: E.N.d.C.L., R.S.A.L., and J.R.C.P. Performed and wrote about tropism analysis: M.A.S. Wrote the final manuscript: E.N.d.C.L., R.S.D., and J.R.C.P. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The Federal University of Sao Paulo, Brazil, Ethics Committee and Human Research Ethics under number #0024/11 approved this study in February 2011, and the Brazilian Ministry of Health authorized it.

HUMAN AND ANIMAL RIGHTS

No animals were used in this research. All human research procedures followed were in accordance with the ethical standards of the committee responsible for human experimentation (institutional and national), and with the Helsinki Declaration of 1975, as revised in 2013.

CONSENT FOR PUBLICATION

Informed consent has not been obtained for this study since HIV genomic sequences were generated for clinical purposes.

AVAILABILITY OF DATA & MATERIALS

The authors confirm that the data supporting the findings of this study are available within the article.

FUNDING

This work was financially supported by Fundação de Amparo a Pesquisa do Estado de Sao Paulo (FAPESP), research grant # 2011/12156-0 to R. S. D. and a PhD scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) to E. N. d. C. L. The funders had no role in study design, data collection or analysis, the decision to publish or the preparation of the manuscript.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Coffin J, Swanstrom R. HIV Pathogenesis: Dynamics and Genetics of Viral Populations and Infected Cells. Cold Spring Harbor Laboratory Press 2013.
- [2] Guha D, Ayyavoo V. Innate immune evasion strategies by human immunodeficiency virus type 1. ISRN AIDS 2013; 2013954806 [http://dx.doi.org/10.1155/2013/954806] [PMID: 24052891]
- [3] Eggink D, de Taeye SW, Bontjer I, *et al.* HIV -1 escape from a peptidic anchor inhibitor by 2 envelope glycoprotein spike stabilization. J Virol 2016. [http://dx.doi.org/10.1128/JVI.01616-16] [PMID: 27654295]
- [4] Merk A, Subramaniam S. HIV-1 envelope glycoprotein structure. Curr Opin Struct Biol 2013; 23(2): 268-76. [http://dx.doi.org/10.1016/j.sbi.2013.03.007] [PMID: 23602427]
- [5] Panos G, Nelson M. HIV-1 tropism. Future Medicine 2007. [http://dx.doi.org/10.2217/17520363.1.4.473]
- [6] Geller R, Domingo-Calap P, Cuevas JM, Rossillo P, Negroni M, Sanjuán R. The external domains of the HIV-1 envelope are a mutational cold spot. Nat Commun 2015; 6: 8571. [http://dx.doi.org/10.1038/ncomms9571] [PMID: 26450412]
- [7] MacNeil A, Sankalé JL, Meloni ST, *et al.* Long-Term inpatient viral evolution during HIV-2 infection JID 2007. [http://dx.doi.org/10.1086/511308]
- [8] Lee HY, Perelson AS, Park SC, Leitner T. Dynamic correlation between intrahost HIV-1 quasispecies evolution and disease progression. PLOS Comput Biol 2008; 4(12):e1000240 [http://dx.doi.org/10.1371/journal.pcbi.1000240] [PMID: 19079613]
- [9] Salemi M. The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. Infect Dis Rep 2013; 5(Suppl. 1):e3 [http://dx.doi.org/10.4081/idr.2013.s1.e3] [PMID: 24470967]
- [10] Mikhail M, Wang B, Lemey P, *et al.* Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort. Retrovirology 2005; 2: 41. [http://dx.doi.org/10.1186/1742-4690-2-41] [PMID: 15985187]
- [11] Borrego P, Marcelino JM, Rocha C, *et al.* The role of the humoral immune response in the molecular evolution of the envelope C2, V3 and C3 regions in chronically HIV-2 infected patients. Retrovirology 2008; 5: 78. [http://dx.doi.org/10.1186/1742-4690-5-78] [PMID: 18778482]
- [12] Yockey HP. Information theory, evolution, and the origin of life. Cambridge University Press 2005. www.krusch.com/books/evolution/Information_Theory_Evolution_Origin_Life_Yockey.pdf [http://dx.doi.org/10.1017/CBO9780511546433]
- [13] Erill I. Information Theory and Biological sequences: insights from an evolutionary perspective 2016. Available in : www.erillab.umbc.edu/files/2016/04/978-1-62100-325-0_ch1.pdf
- [14] Pan K, Deem MW. Quantifying selection and diversity in viruses by entropy methods, with application to the haemagglutinin of H3N2 influenza. J R Soc Interface 2011; 8(64): 1644-53. [http://dx.doi.org/10.1098/rsif.2011.0105] [PMID: 21543352]
- [15] Patronovand A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. 2013; 3: p. (1)120139. [http://dx.doi.org/10.1098/rsob.120139]
- [16] Liu MKP, Hawkins N, Ritchie AJ, *et al.* Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. J Clin Invest 2013; 123(1): 380-93. [http://dx.doi.org/10.1172/JCI65330] [PMID: 23221345]
- [17] Hancock G, Yang H, Yorke E, *et al.* Identification of effective subdominant anti-HIV-1 CD8+ T cells within entire post-infection and post-vaccination immune responses. PLoS Pathog 2015; 11(2):e1004658 [http://dx.doi.org/10.1371/journal.ppat.1004658] [PMID: 25723536]
- [18] Adami C. The use of information theory in evolutionary biology. Ann N Y Acad Sci 2012; 1256: 49-65. [http://dx.doi.org/10.1111/j.1749-6632.2011.06422.x] [PMID: 22320231]
- [19] Diaz RS, Zhang L, Busch MP, *et al.* Divergence of HIV-1 quasispecies in an epidemiologic cluster. AIDS 1997; 11(4): 415-22.
- [20] Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23(21): 2947-8. [http://dx.doi.org/10.1093/bioinformatics/btm404] [PMID: 17846036]
- [21] Gatenby RA, Frieden BR. Information theory in living systems, methods, applications, and challenges. Bull Math Biol 2007; 69(2): 635-57. [http://dx.doi.org/10.1007/s11538-006-9141-5] [PMID: 17083004]
- [22] MathWorks. Bioinformatics Toolbox™: User's Guide (R2017b) Available from www.mathworks.com/help/pdf_doc/bioinfo/bioinfo_ug.pdf
- [23] Chen GW, Chang SC, Mok CK, *et al.* Genomic signatures of human versus avian influenza A viruses. Emerg Infect Dis 2006; 12(9): 1353-60. [http://dx.doi.org/10.3201/eid1209.060276] [PMID: 17073083]
- [24] Impact of antiretroviral resistance and virologic failure on HIV-1 informational entropy. J Antimicrob Chemother 2018. [http://dx.doi.org/10.1093/jac/dkx50]
- [25] Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G >A hypermutation. Bioinformatics 2000; 16(4): 400-1.
- [26] Beerenwinkel N, Däumer M, Oette M, *et al.* Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res 2003; 31(13): 3850-5. [http://dx.doi.org/10.1093/nar/gkg575] [PMID: 12824435]
- [27] Vandekerckhove LP, Wensing AM, Kaiser R, *et al.* European guidelines on the clinical management of HIV-1 tropism testing. Lancet Infect Dis 2011; 11(5): 394-407. [http://dx.doi.org/10.1016/S1473-3099(10)70319-4] [PMID: 21429803]
- [28] Motulsky HJ. GraphPad Statistics Guide Available from http://www.graphpad.com/guides/prism/7/statistics/index.htm
- [29] BioAfrica Bioinformatics & Genomics For Health and Life Sciences in Africa GP 120 – ENV Glycoprotein 120 Available from : http://www.bioafrica.net/proteomics/ENV-GP120prot.html
- [30] Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 1998; 393(6686): 648-59. [http://dx.doi.org/10.1038/31405] [PMID: 9641677]
- [31] Kwong PD, Doyle ML, Casper DJ, *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature 2002; 420(6916): 678-82. [http://dx.doi.org/10.1038/nature01188] [PMID: 12478295]
- [32] Koch M, Pancera M, Kwong PD, *et al.* Structure-based, targeted deglycosylation of HIV-1 gp120 and effects on neutralization sensitivity and antibody recognition. Virology 2003; 313(2): 387-400. [http://dx.doi.org/10.1016/S0042-6822(03)00294-0] [PMID: 12954207]
- [33] Pancera M, Majeed S. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility In: PNAS. 2010. [http://dx.doi.org/10.1073/pnas.0911004107]
- [34] Farzan M, Mirzabekov T, Kolchinsky P, *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. Cell 1999; 96: 667-76.
- [35] Hartley O, Klasse PJ, Sattentau QJ, *et al.* V3: HIV's switch-hitter AIDS Res Hum Retroviruses 2005; 21(2): 171-89.
- [36] Sucupira MC, Sanabani S, Cortes RM, *et al.* Faster HIV-1 disease progression among Brazilian individuals recently infected with CXCR4-utilizing strains. PLoS One 2012; 7(1):e30292 [http://dx.doi.org/10.1371/journal.pone.0030292] [PMID: 22291931]
- [37] Arif MS, Hunter J, Léda AR, *et al.* Pace of Coreceptor Tropism Switch in HIV-1-Infected Individuals after Recent Infection. J Virol 2017; 91(19):e00793-17 [http://dx.doi.org/10.1128/JVI.00793-17] [PMID: 28659473]
- [38] Hayashida T, Tsuchiya K, Kikuchi Y, Oka S, Gatanaga H. Emergence of CXCR4-tropic HIV-1 variants followed by rapid disease progression in hemophiliac slow progressors. PLoS One 2017; 12(5):e0177033 [http://dx.doi.org/10.1371/journal.pone.0177033] [PMID: 28472121]
- [39] Montagna C, De Crignis E, Bon I, *et al.* V3 net charge: additional tool in HIV-1 tropism prediction. AIDS Res Hum Retroviruses 2014; 30(12): 1203-12. [http://dx.doi.org/10.1089/aid.2014.0006] [PMID: 25322170]
- [40] Shen HS, Yin J, Leng F, *et al.* HIV coreceptor tropism determination and mutational pattern identification. Sci Rep 2016; 6: 21280. [http://dx.doi.org/10.1038/srep21280] [PMID: 26883082]
- [41] Yu B, Fonseca DPAJ, O'Rourke SM, Berman PW. Protease cleavage sites in HIV-1 gp120 recognized by antigen processing enzymes are

- conserved and located at receptor binding sites. *J Virol* 2010; 84(3): 1513-26.
[http://dx.doi.org/10.1128/JVI.01765-09] [PMID: 19939935]
- [42] West AP Jr, Scharf L, Scheid JF, Klein F, Bjorkman PJ, Nussenzweig MC. Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* 2014; 156(4): 633-48.
[http://dx.doi.org/10.1016/j.cell.2014.01.052] [PMID: 24529371]
- [43] Hamoudi M, Simon-Loriere E, Gasser R, Negroni M. Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. *Retrovirology* 2013; 10: 114.
[http://dx.doi.org/10.1186/1742-4690-10-114] [PMID: 24156625]

© 2020 de Carvalho Lima *et al.*

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.