

Characterization of Variable Regions of the Gp120 Protein from HIV-1 Subtype C Virus Variants Obtained from Individuals at Different Disease Stages in Sub-Saharan Africa

A. Cenci¹, L. Tavoschi¹, G. D' Avenio², P. Narino¹, S. Becattini^{1,6}, D. Bernasconi^{1,7}, M. Chiappi^{1,8}, L. La Torre¹, H. Sukati³, E. Vardas⁴, A. Lo Presti⁵, E. Cella⁵, M. Ciccozzi⁵, O. Picconi¹, P. Monini¹, B. Ensoli¹ and S. Buttò^{1*}

¹Istituto Superiore di Sanità, National AIDS Center, Rome, Italy

²Istituto Superiore di Sanità, Department of Technology and Health, Rome, Italy

³Ministry of Health, National Center Public Health Laboratory, Manzini, Swaziland

⁴Stellenbosch University, Division of Medical Virology, Stellenbosch, South Africa

⁵Istituto Superiore di Sanità, Department of Infectious, Parasitic and Immunomediated Diseases, Rome, Italy

⁶Institute for Research in Biomedicine, Bellinzona, Switzerland

⁷Abbott diagnostics Italy

⁸Departamento de Estructura de Macromoléculas, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid, Spain

Abstract

Background: The development of a vaccine against HIV/AIDS capable of preventing virus infection has been hampered by the HIV envelope (Env) heterogeneity that makes it difficult to induce neutralizing antibodies against Env proteins from different HIV clades. Several studies have indicated that gp120 Env protein sequence tends to change considerably during the course of HIV disease which allows the virus to escape the immune responses. In order to define gp120 sequence changes, we have characterized the V1, V2, V4 and V5 variable regions of gp120 variants from 72 HIV-1-clade-C-infected subjects from South Africa and Swaziland, which were naïve to antiretroviral (ARV) therapy and at different disease stages. Sequence characteristics, such as amino acid sequence length, presence of Putative N-Glycosylation Sites (PNGSs) and electric charge were investigated.

Methods: According to the Avidity Index value and CD4+ T cell count, patients were classified for disease stage in three groups: recent, chronic and late stage, each one comprised of 24 patients. The V1 to V5 Env variable regions were directly PCR amplified from plasma virus RNA and sequenced.

Results: A significant increase in the amino acid sequence length of V1 and V4 domains, and a corresponding increase of the "shifting" PNGSs were observed in the HIV variants obtained from individuals at chronic stage of disease, as compared to the recent infection group. Finally, a significant increase of the net electric positive charge of the V5 loop was found in the HIV variants from the group of subjects with late disease, as compared to the chronic disease group.

Conclusion: We conclude that changes in sequence length, glycosylation pattern and net electrical charge in the variable V1, V4 and V5 regions of gp120 occur in the course of HIV infection, possibly in response to the pressure of the host immune response.

Keywords: HIV-1 Subtype C; Envelope

Introduction

According to 2011 UNAIDS statistics, the world's AIDS epidemic has hit a plateau, with 2.7 million people becoming newly infected each year for the last five years [1]. Two-thirds of all people infected with HIV live in sub-Saharan Africa with severe demographic effects and country development impact. In this geographical area, South Africa and Swaziland are among the countries most affected by HIV/AIDS. UNAIDS has estimated that in 2009 the total number of persons living with HIV in South Africa was 5.7 million, a number that makes South Africa one of the countries with the largest HIV epidemic in the world [2]. Recent data also indicate that in South Africa incidence among men and women aged 15-49 was estimated to be 1.3/100/pyar in the period 2005-2008 [3]. For Swaziland, the 26% HIV prevalence in the adult population found in 2006 through a national population-based survey, is the highest ever documented [4]. Finally, South Africa and Swaziland are also among the countries in the world, where the number of recent infections is extremely high [5].

The development of a vaccine capable of preventing HIV infection and reducing virus transmission and spread would be essential to reduce the burden of AIDS and AIDS-related diseases, in particular in

developing countries. However, such a preventative vaccine still remains a formidable challenge. One of the obstacles to its development is the design of an immunogen that elicits potent and broadly neutralizing antibodies, capable of preventing HIV infection. The primary targets of this neutralising response are epitopes in the HIV envelope (Env) that are involved in infectivity and transmission. However, attempts to induce these types of neutralising antibodies have failed so far [6], due to the high variability of the Env proteins, from different clades and in the same subject, during the disease that allows the virus to escape the host immune response [7].

***Corresponding author:** Stefano Buttò, National AIDS Center, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Rome, Italy, Tel: +39-06-49903249; Fax: +39-06-49903002; E-mail: stefano.butto@iss.it

Received April 16, 2012; Accepted May 16, 2012; Published May 20, 2012

Citation: Cenci A, Tavoschi L, D' Avenio G, Narino P, Becattini S, et al. (2012) Characterization of Variable Regions of the Gp120 Protein from HIV-1 Subtype C Virus Variants Obtained from Individuals at Different Disease Stages in Sub-Saharan Africa. J AIDS Clinic Res S8:006. doi:10.4172/2155-6113.S8-006

Copyright: © 2012 Cenci A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Genetic diversity of HIV-1 is represented by distinct subtypes or clades, A to K, as well as by recombinants [8]. The HIV-1 subtype C is prevalent in Sub-Saharan Africa and is widespread worldwide, being responsible of more than 52% of total HIV-1 infections [9,10].

Five loop regions of the gp120, V1 to V5, are particularly variable and play a major role in infectivity, transmission and resistance to neutralization [11-13]. Early data on HIV-1 clade B, and more recent data on clade C, have shown that the V1 region is highly variable, protects the CD4 binding-site from neutralising antibodies and may play a role in infectivity and tropism [14,15]. The V2 region, closely interacting with V1, is less variable, can bind the Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) [16], and also plays a role in infectivity and tropism [16,17]. The V1 and V2 regions, together, have been shown to modulate neutralization sensitivity by hiding conserved epitopes, such as the coreceptor binding site [18-20]. The V3 region has been shown to bind both CCR5 and CXCR4 cellular co-receptors [21,22]. In HIV-1 subtype C viruses this region is more conserved than in other clades [23,24], and preferentially binds the CCR5 co-receptor, independently of the disease stage [25,26]. Finally, the function of V4 and V5 regions has not been clearly identified, although they are likely to influence Env conformation and glycan packing [27,28], thereby limiting accessibility to neutralization determinants by steric hindrance. In addition, some data indicate that epitopes in the C3-V4 region are major targets of the early autologous neutralizing response in HIV-1 subtype C infection [29].

HIV-1 Env gp120 is among the most heavily glycosylated proteins in nature [30-32]. Alterations of the glycosylation pattern can impact protein folding [33] and play a role in masking linear and conformational epitopes that are sensitive targets for neutralization [34]. Changes in the glycosylation pattern may also affect and influence co-receptor binding, thereby playing a role in viral tropism [35]. It has also been demonstrated that N-glycans in gp120 may play a role in viral transmission [11] and infectivity through the interaction between high mannose oligosaccharides and the DC-SIGN or other C-type lectins [36]. Altered patterns of glycosylation can modulate the immunogenic properties of the gp120 and contribute to escape from humoral and T-cell responses [14,37], thus providing a "shield" against the immune response of the host [27].

Increasing evidence, based on restricted numbers of donor/recipient couples and individuals with acute infection, suggests that a few or just one single viral variant, characterized by a more compact gp120 Env protein, can establish the initial infection [11,38-41]. More recently, it has been demonstrated that primary infection may be mediated by viruses with relatively short variable regions of gp120 and a low number of PNGSs and that these variants show a higher sensitivity to neutralization by sera [38,42,43]. Other data indicate that variants from different HIV-1 clades, isolated from recently infected patients, show shorter V1-V2 sequences than variants isolated from chronically infected individuals, and that these variants are also more sensitive to neutralisation by patient sera than those isolated during chronic infection [44-46]. Finally, it has been recently reported that variations in the net electric charge in both variable and constant regions of the virus may be responsible of virus resistance to neutralisation [23,34,47]. Therefore, it is plausible that during the course of the disease the sequence characteristics of Env variable domains may have an impact on the selection of virus variants and on their ability to be transmitted and/or escape immune control.

In order to identify key putative features of the Env sequence, we have performed a cross-sectional study on 72 HIV-1 clade C-infected

individuals from Swaziland and South Africa at different disease stages, from whom we obtained virus variants that have been characterised for amino acid sequence length, glycosylation pattern and net electric charge of their gp120 variable regions.

Materials and Methods

Sample collection

A total of 72 plasma samples were collected in the period 2005-2007 from HIV-positive individuals, living in the Southern Africa region, all naïve for Antiretroviral Therapy (ART). Samples were obtained from 24 individuals enrolled during the 2006 Swaziland HIV National Serosurvey [4], 24 individuals attending the Chris Hani Baragwanath Hospital (CHBH) in Soweto, Johannesburg, South-Africa, in the framework of the activities included in the AVIP project (www.avip-eu.org) and 24 individuals attending the HIV/AIDS National Referral Laboratory (NRL) Hospital in Mbabane, Swaziland, in the framework of activities of projects from the Italian AIDS Programme. Ethical clearance for these studies was previously obtained from local Ethical Committees. Disease stage of each enrolled individual was determined on the basis of the Avidity Index [48,49] and CD4+ T cell count, as described in the Results section.

CD4+ T cell count

CD4+ T cell count was performed by means of MultiTEST and TruCOUNT tubes (Beckton Dickinson) according to the manufacturer's instructions.

Avidity Index (AI) assay

Avidity Index Assay was performed as previously described [48,49]. An Index ≤ 0.8 identified an HIV infection occurred within 6 months from blood drawing (here defined as recent infection) [48,49].

Viral RNA extraction and sequence amplification

Viral RNA was extracted from 0.5 ml of plasma using Qiamp viral RNA miniKit (Qiagen), after treatment with Heparinase (Sigma). The V1-V5 coding region in the Env gene was amplified by RT-PCR using SuperScript™ One-Step RT-PCR System (Invitrogen) and specifically designed external primers for clade C HIV-1. Primers were: AC-Env Outer For 5'-CAGATGCATGAGGATATAATCA-3'; ED12m Outer Rev 5'-AGTGCTTCCTTGCTGCTCCCAA3'.

The RT-PCR mix was composed of 0.5 to 1 μ g of RNA that were added to an RT/Taq buffer mixture containing 0.4 mM dATP, dCTP, dTTP, dGTP (Roche), 2.4 mM MgSO₄ (Invitrogen), 1 μ l RT/Taq-platinum (Invitrogen), 40 U/ μ l of RNase inhibitor (Invitrogen), and 10 μ M primers (MWB-Biotech). Reaction was carried out in a thermal cycler (Eppendorf) according to the following program: 45°C for 30', for retrotranscription and 94°C for 2' for the RT denaturation; the resulting cDNA was amplified as follows: 94°C for 15", 50°C for 30" and 68°C for 1'30" for a total of 40 cycles. A final extension step followed for 7' at 68°C.

The resulting PCR product was again amplified in a nested PCR, using specifically designed primers identified in conserved external portions enclosing each target nucleotide sequence (i.e. V1-V2, V3-C3 and V4-C4-V5 Env regions). Primers specific for each region are reported in Table 1.

Each nested reaction was conducted as follows: an aliquot of 1-5 μ l of the amplified product was added to a reaction mixture containing 200 μ M dNTPS (Roche), 2.5 mM MgCl₂ (Invitrogen), the couple of

ENV Region	Primers I	Primers II
V1-V2	ForA 5'ACCCCACTCTGTGCACTTT3' RevA 5'TATTACACTTTAGAATCGCA3'	ForB 5'AAGTTGACCCCACTCTGTGT3' RevB 5'CTTTAGAATCGCATAACCAG 3'
V3-C3	ForA 5'CTGTAAATGGTAGCCTAGC3' RevA 5'GCAATAGAAAAATTCCTCC3'	ForB 5'CACAGTACAATGTACACATG3' RevB 5'RCAATAGAAAAATTCCTC3'
V4-C4-V5	ForA 5'GTRGAGGAGAATTTTCTATTG3' RevA 5'TATAATTCACCTCTCCARTTGTC3'	ForB 5' TTTAATTGTRGAGGAGAATTTTCTATTG3' RevB 5' TATTTATAATTCACCTCTCCAATTGTC3'

In case amplification failed using the first couple of primers (Primers I), nested PCR was repeated using another couple of inner primers (Primers II).

Table 1: Primers used for nested PCR amplification of selected regions of gp120 HIV *env*.

20 μ M primers corresponding to the region that had to be amplified (MWB-Biotech), 2.5 U/ μ l of AmpliTaq Gold DNA polymerase and 1x AmpliTaq Buffer (Applied Biosystems). Amplifications were carried out specifically for each region as follows: V1-V2: 96°C for 7', then 15" at 94°C, 30" at 48°C, 30" at 72°C (for 40 cycles); V3-C3: 96°C for 7', then 15" at 94°C, 30" at 50°C, 30" at 72°C (for 40 cycles); V4-C4-V5: 96°C for 7', then 15" at 94°C, 30" at 44°C, 30" at 72°C (for 40 cycles).

All these PCR reactions were followed by a final extension step of 7' at 68°C.

DNA purification and sequencing

Amplified DNA of each region was purified using Qiaquick PCR purification kit (Qiagen), according to the manufacturer's protocol. The DNA samples were then quantified and checked for purity by measuring absorbance at 260 nm and 280 nm to estimate contaminants.

The V1-V2, V3-C3 and V4-C4-V5 amplified nucleotide sequences were then sequenced according to the Sanger method [50] using the same primers as nested PCR.

Sequence analysis

Sequencing was performed on uncloned PCR products to identify the prevalent viral quasispecies. The electropherogram was edited using Chromas Pro (www.techesium.com.au/ChromasPro.html). All sequences were aligned using Clustal X [51] and corrected for multiple alignment by manual editing.

The nucleotide sequences were translated into aminoacid sequences using GeneRunner (www.generunner.net) and further codon-aligned using BioEdit [52]. The V1, V2 and V1-V2 regions were defined as sequences comprised between the corresponding HXBc2 envelope aminoacids 131 to 157, 158 to 196 and 131 to 196, respectively (Accession number AAB50262-1, Geneprot). The V3 region was defined as the sequence comprised between the corresponding HXBc2 envelope aminoacids 296 to 331. The V4 and V5 regions were defined as sequences comprised between the corresponding HXBc2 envelope aminoacids 385 to 418 and 460 to 471, respectively.

The identification of PNGSs in V1, V2, V3, V4 and V5 regions was performed by using N-GLYCOSYTE (<http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>). The analysis of predicted co-receptor usage on V3 region was performed by using Geno2pheno (<http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>).

Phylogenetic analysis was carried out using the PAUP software (version 4.0) [53] with the K81 model of substitution and by the use of both Neighbor-Joining (NJ) and Maximum Likelihood (ML) tree-building methods. The evolutionary model was chosen as the best-fitting nucleotide substitution model in accordance with the results of the Hierarchical Likelihood Ratio Test (HLRT) implemented in MODELTEST software (version 3.6) [54].

The parameters for the nucleotide substitution model were

estimated by the ML method using a NJ tree (Jukes-Cantor distance) as the base tree [55].

The statistical robustness and reliability of the branching order within phylogenetic trees were confirmed by either a bootstrap analysis using 1000 replicates, for the NJ tree, or the zero branch length tests for the ML tree. All calculations were performed with PAUP software (version 4.0) [53].

In order to evaluate genetic distances among sequences from variants isolated during the three disease stages, the Mega 4 program (www.megasoftware.net) with Kimura 2 model was used.

HIV group M subtypes and CRFs sequences were available at the Los Alamos database (www.hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.html).

Statistical analysis

The Wilcoxon rank sum test was used to compare the differences between the groups of recently infected individuals, patients with chronic disease and patients at late disease stage, for each one of the parameters analysed in the study.

Furthermore, to measure existing correlation between sequence length and glycosylation status, a linear regression, using Spearman's correlation test, was performed (Stata 8.2).

For the electric charge analysis of each region, the algebraic sum of all charged residues, both positive and negative, was considered as the total electric charge (Q_{tot}), whereas the sums of all the positive charged residues (Arg, Lys, Hys) and of all negative charged residues (Glu, Asp) were considered as total positive (Q_{pos}) and total negative (Q_{neg}) charge, respectively. An original Matlab program was written to yield the distribution of Q_{tot}, Q_{pos} and Q_{neg} directly from the *fas files with the aligned sequences. Such distributions were analyzed with Wilcoxon's rank sum test, after a preliminary checking of the non-Gaussian nature of the distributions using the Kolmogorov-Smirnov test, as previously described [56].

Results

The 72 HIV-1-infected patients from Swaziland and South Africa, all naïve for antiretroviral therapy, were classified for the disease stage in three groups: Recent Stage (RS), Chronic Stage (CS), and Late Stage (LS), on the basis of the AI assay and the CD4+T cell count. An AI \leq 0.80 identified samples from recently infected individuals with stage (RS), [48,49], whereas patients with chronic and late disease (CS and LS, respectively) were identified on the basis of an AI > 0.80 and CD4+ T cell counts. Specifically, patients with CD4+ T cell count \leq 200/ μ l were considered to be at a late stage of disease, whereas the group of patients with chronic disease was identified by CD4+ T cell counts > 200/ μ l (Table 2).

A Maximum Likelihood (ML) phylogenetic tree, based on the V3-V5 coding region, classified all variants obtained from the patients

within the HIV-1 clade C subtype (Figure 1). The tree also showed a mixing of the Env sequences topology, independently of the patient region of origin. Moreover, nucleotide distances, implemented by Mega software, between viruses isolated from patients in Swaziland and viruses isolated from patients in South Africa, never exceeded 17%, confirming a close relationship among the isolates from the two countries.

Amino acid sequence alignment of the V1-V2, V3, and V4-V5 regions showed high sequence variability in V1 and V4 sequences, for both amino acid substitutions and insertion/deletions of short amino acid sequences, mainly clustering in the C-terminal portion of both regions. The V2 and V5 regions were less variable, whereas the V3 region appeared well conserved among all the isolates. The high sequence variability originated also new PNGSs that clustered in the C

terminal portion of the V1, V2 and V4 loops, in particular in the group of patients with chronic disease.

The V3 region was extremely conserved for both amino acid sequence length and PNGSs number (median amino acid sequence length: 35aa; median PNGS number: 1) in all three disease stages, as previously described for HIV-1 subtype C virus [23] (data not shown). Therefore, due to its high degree of sequence conservation, the V3 region was not included in the subsequent analyses of the characteristics of the gp120 variable sequences.

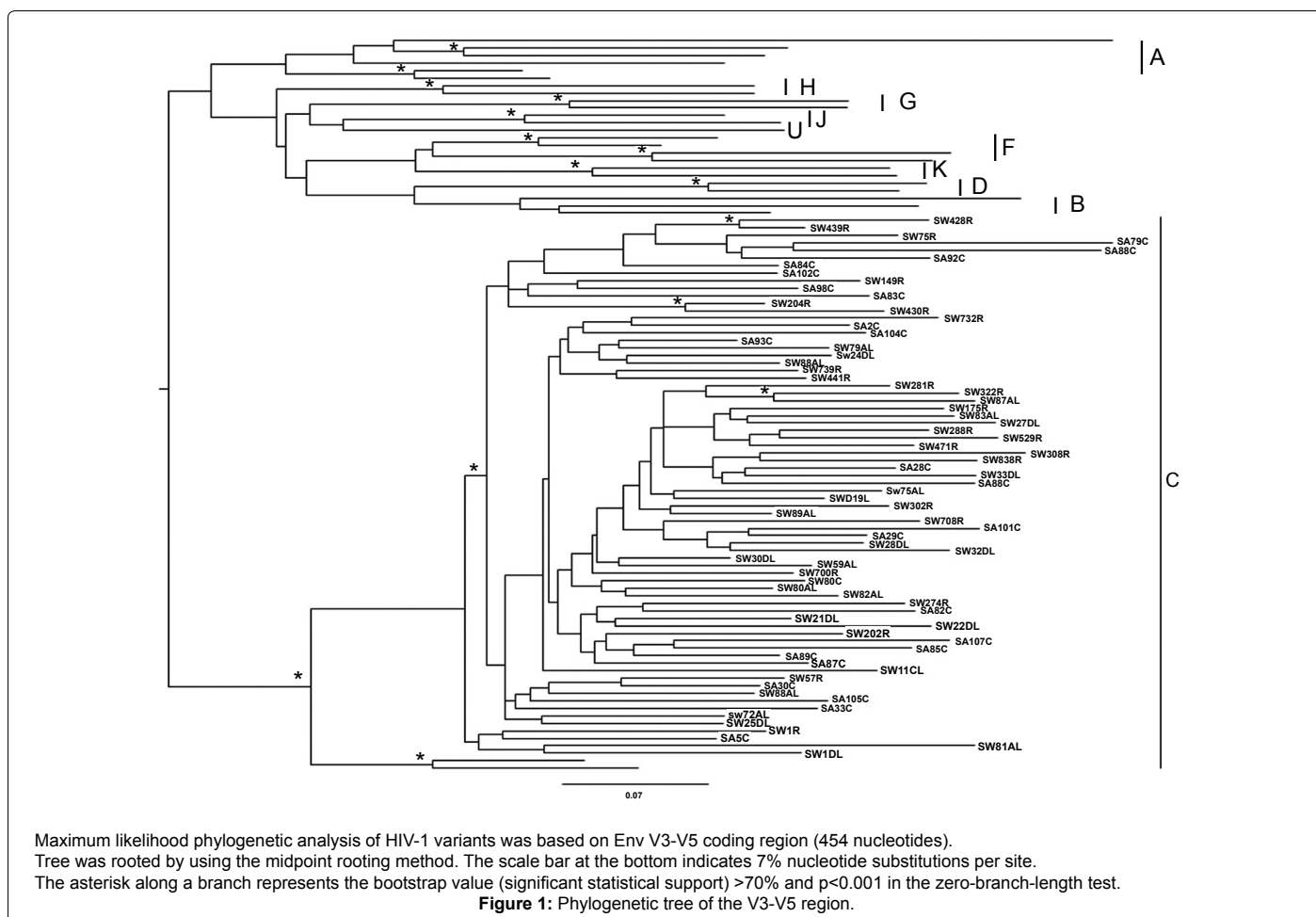
A comparative box plot analysis of amino acid sequence length and PNGS number for each variable region is reported in figures 2 and 3, respectively. Amino acid sequence length of the V1 and V4 regions significantly increased in the CS patients when compared to that observed in patients with recent infection (RS) (p values of 0.0083 and 0.0158, respectively) (figure 2). In addition, there was a statistically significant increase of the length of the V4 region in the group of LS patients compared to the RS patients group (p=0.0208). A reverse trend to shorter sequences (not statistically significant) was observed in LS individuals for both V1 and V4 regions. The V2 domain showed a somewhat inverted behaviour for length variability, although not statistically significant, since the sequence length tended to be slightly reduced in patients in the chronic disease stage when compared to individuals in the early and late disease stages. Finally, no statistically significant changes of amino acid length of the V5 region were present in each of the three groups of patients.

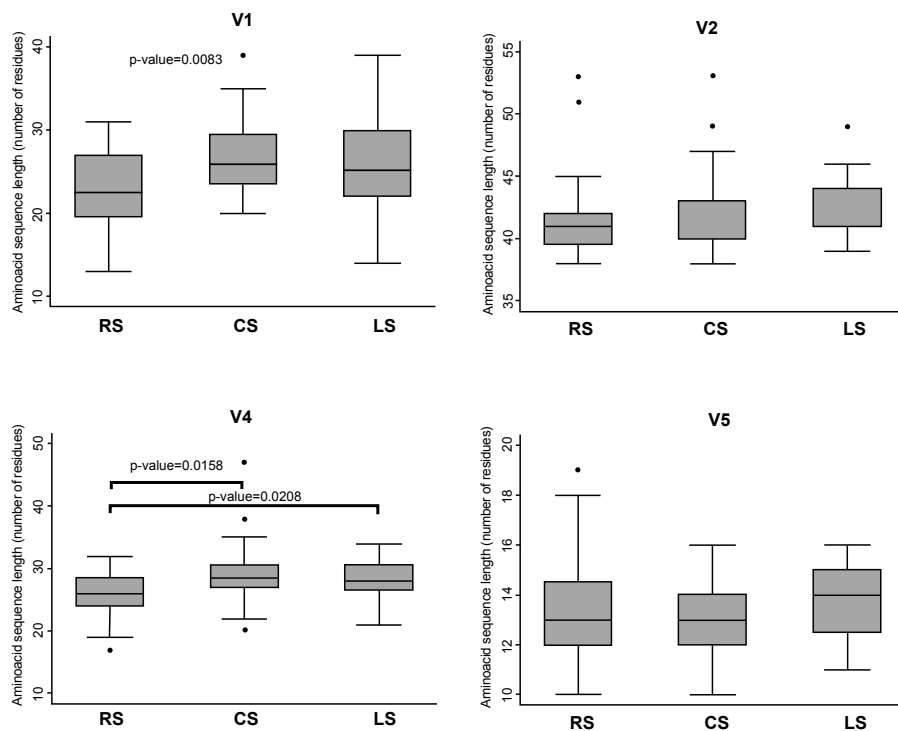
	Number of patients	A I	CD4+ T-cell count Range; [median](cell/μl)
RS	24	≤0.80	NA*
CS	24	>0.80	418–881; [475]
LS	24	>0.80	17-199; [107.5]

*NA: Not Applicable.

RS= Recent disease Stage; CS= Chronic disease Stage; LS= Late disease Stage.

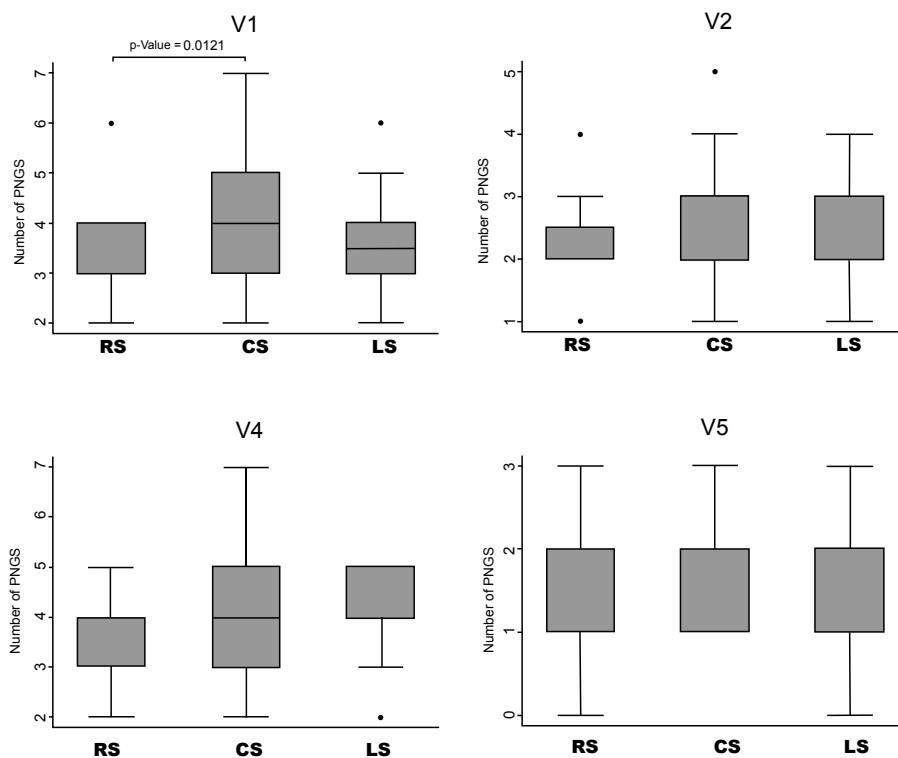
Table 2: Classification of patients for disease stage according to AI and CD4+ T cell counts.





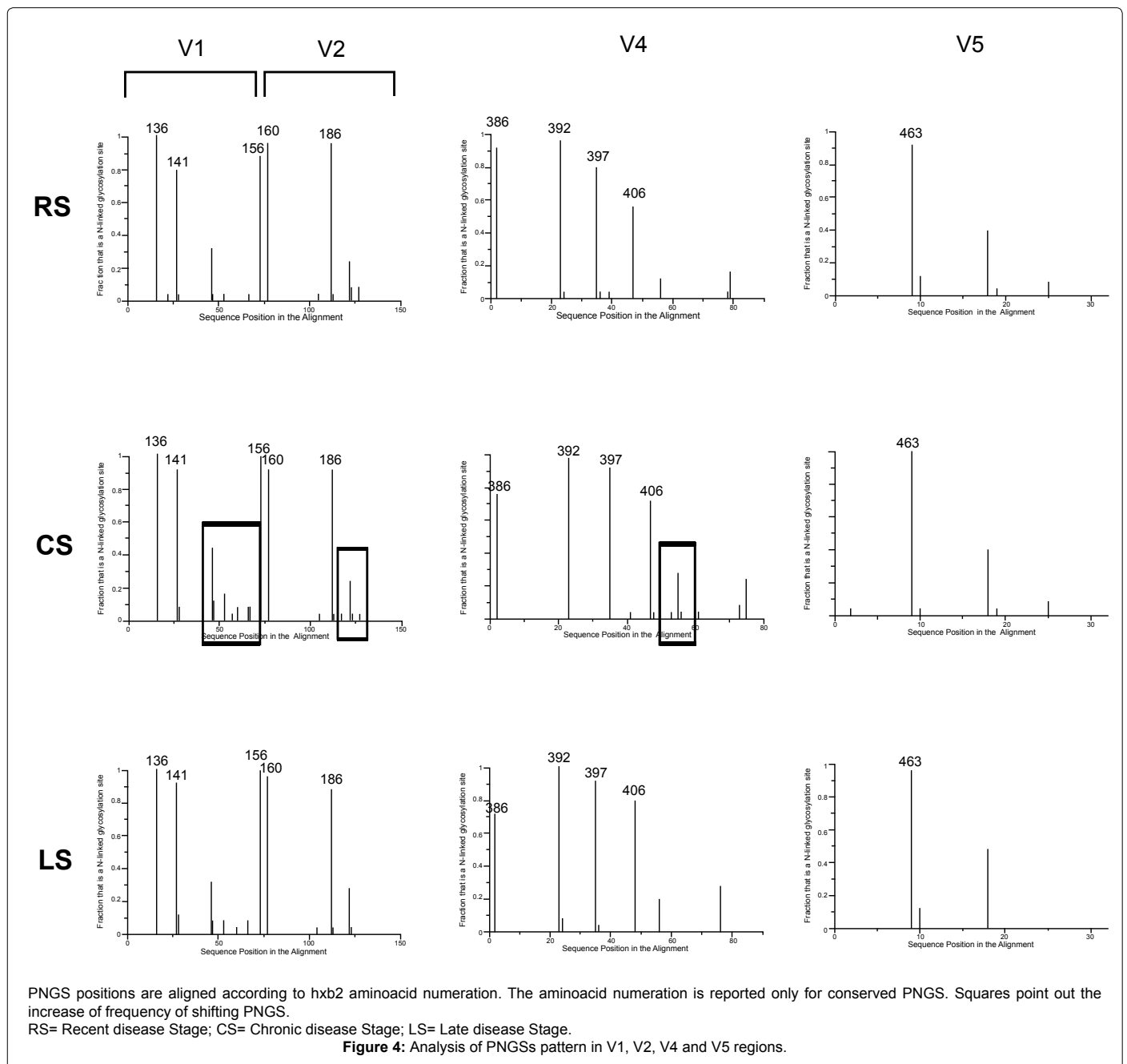
Significant p values are shown. The box plots denote the median, minimum and maximum values.
RS= Recent disease Stage; CS= Chronic disease Stage; LS= Late disease Stage.

Figure 2: Box plot analysis of amino acid sequence length of V1, V2, V4 and V5 regions from patients at different disease stages.



Significant p values are shown. The box plots denote the median, minimum and maximum values.
RS= Recent disease Stage; CS= Chronic disease Stage; LS= Late disease Stage.

Figure 3: Box plot analysis of PNGSs number in V1, V2, V4 and V5 regions from patients at different disease stages



The box plot analysis of PNGSs number is reported in figure 3. The number of PNGSs in the V1 region was significantly increased in CS patients, as compared to the RS patients group ($p=0.0121$). Furthermore, a slight not statistically significant, reduction of PNGSs number in V1 during the late stage of disease was observed. In addition, a slight not statistically significant increase of PNGS number was observed for the V4 region in CS patients, compared to the RS patients group, while the number remained quite constant in the late infection stage. Neither statistically significant changes in the PNGSs number, nor trends, were present in the V2 and V5 regions when the three groups of patients were compared.

A quantitative analysis of the PNGSs distribution revealed a stable presence of highly conserved PNGS, with an expression frequency

≥ 0.70 , irrespective of the disease stage, in the positions N136, N141, N156, N160, N186, N301, N386, N392, N397, and N463 of the corresponding HXBc2 sequence, distributed in all the five variable regions.

The qualitative analysis of the distribution of PNGSs among the different groups of patients is reported in figure 4. Besides the presence of the conserved PNGSs, a number of PNGSs with a lower expression level could be identified (referred as shifting PNGSs) [31]. These PNGS clustered in the C-terminal portion of V1 and V4 and, more moderately, in V2, in the CS group of patients, as compared to RS patients. Their mean expression frequency was < 0.40 and only a few of them were still present in the group of patients in the late stage of disease. An exception was represented by the PNGS N406 in the V4 region that had

an intermediate behaviour, being expressed with a frequency <0.5 in the group of recently infected patients and increasing up to >0.7 in the groups of patients in the chronic and late disease stages.

To investigate if the variability patterns of amino acid length and PNGSs distribution were related, a regression analysis was carried out for each variable V1, V2, V4 and V5 domain, individually. The results indicated that a strong statistically significant positive correlation exists between length and PNGSs number, associated with positive regression coefficients.

To complete the analysis of sequence characteristics, we have calculated the net electric charge of each region, considering positive, negative and total electric charge distributions. Total electric charge (Q_{tot}) of V1, V2 and V4 regions was found to be stable, the transition between different stages of disease being characterized by no significant changes. In the V5 region, the Q_{tot} was lower in the CS patients disease, as compared to the groups of patients at both recent and late disease stage (figure 5), although the change was not statistically significant. However, when positive (Q_{pos}) and negative (Q_{neg}) electric charges were considered separately, a statistically significant increase of Q_{pos} of V5 region was observed in the LS patients group, compared to the CS patients ($p=0.027$), whereas Q_{neg} was similar in all the three groups (figure 5).

Discussion

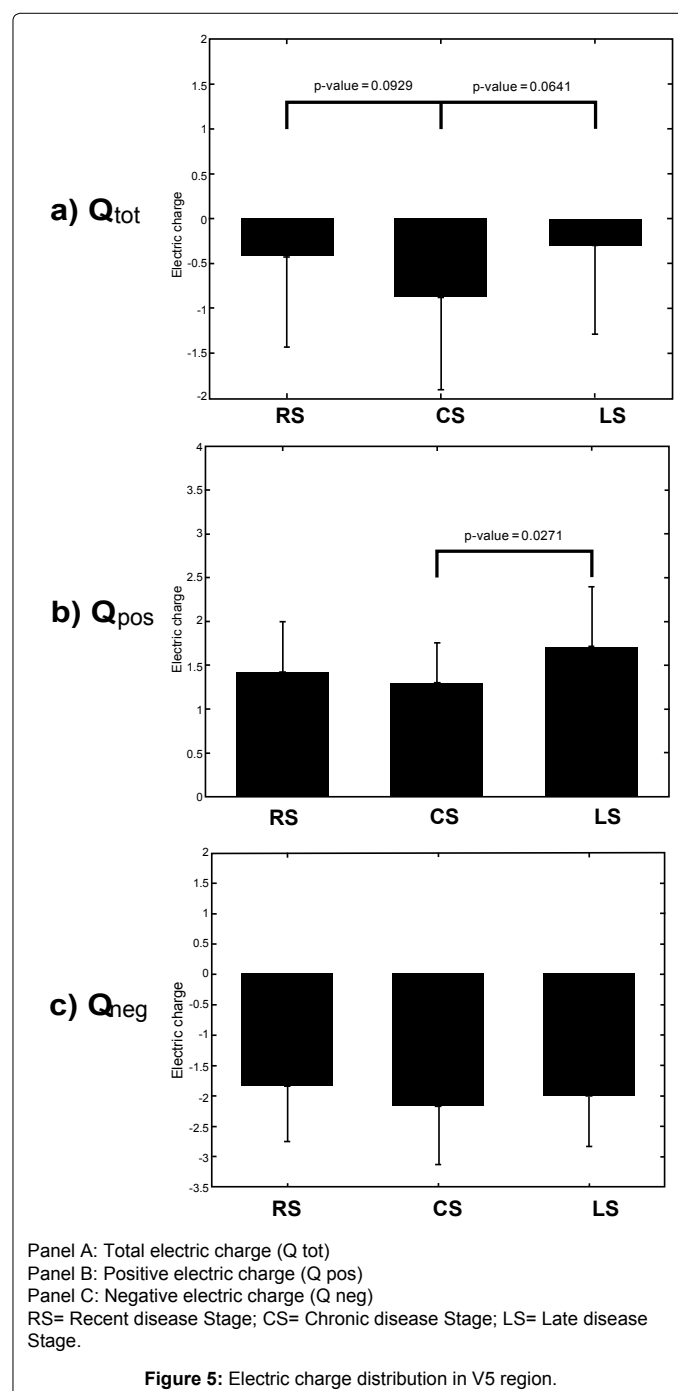
We have conducted a cross-sectional study in patients at different disease stages in South Africa and Swaziland, where HIV-1 subtype C virus is predominant [57-59], with the aim to investigate changes of sequence characteristics of gp120 variable domains from virus variants obtained during the course of the disease. In order to avoid bias due to the pressure of antiretroviral drugs, we included in the study only patients naive for antiretroviral therapy. In addition, we excluded sampling bias due to collection of samples from two different countries, although from the same geographical region, by generating a Maximum Likelihood phylogenetic tree which indicated that the two subgroups were similar (nucleotide distance $<10\%$).

Our data on sequence analysis of gp120 variable domains confirmed an increase in overall amino acid variability in the chronic stage of the disease, predominantly in the V1 and V4 regions and, to a lower extent, in the V2 and V5 regions with the exception of the V3 region that was, instead, highly conserved, in all the three disease stages, in line with previously published studies [60].

The V1 and V4 sequence lengths significantly increased in the group of patients in the chronic disease stage, compared to those in the early stage. These data, among the few ones obtained in cross-sectional studies, support what was previously observed in longitudinal on HIV-1 C and non-C subtypes, namely that HIV-1 isolates from patients with chronic disease are prone to length increase of the variable regions of gp120 which promotes resistance to neutralization [11,38,43,45,61-63]. It is possible that these large V1/V2 domains contain epitopes recognized by neutralizing antibodies, in addition to those present in the V4 region of the HIV-1 C subtype [29]. This is further confirmed by recent data obtained on clade C variants indicating that shorter V1 loops can expose the epitope recognized by anti-subtype C antibodies directed to the V1/V2 region, resulting in higher neutralization titres [64].

Our data also show a slight decrease of sequence length in the V1 and V4 loops, during the late stage of the disease when compared to the chronic stage. We could speculate that sequence length reduction

is a consequence of the reduced pressure of the immune response in late disease stage. In fact, variants bearing shorter V1 and V4 sequences may be favoured in late stages of the disease, as is the case with recent infections, when the HIV-specific immune response has not completely developed. This hypothesis is further supported by published data on subtype B, showing that the reduction of immunological pressure in the late disease stage may allow the re-emergence of previous existing variants [63]. In addition, a CD4+ T cell count cut-off of 150 cells/ μ l has been proposed to define the drop down of the selective pressure on Env variable domains [65], which is consistent with the CD4+ T cell count median value of 107.5 cells/ μ l in our patients in the late disease stage.



The PNGSs distribution in the V1, V2, V4 and V5 domains during the course of the disease followed a pattern comparable to the one described for the sequences length. The existence of a positive correlation between length and PNGS number suggests that for the sequence length there could be a viral mechanism of evolution towards variants with a more glycosylated Env protein variants.

The distribution of conserved and shifting PNGS over the variable loops sequences also revealed some other intriguing possibilities. The analysis confirmed the presence of conserved PNGSs characterized by a steady high expression level (>70% throughout the disease stages), as previously described [66]. These PNGS are likely to be subjected to a negative selective pressure by the immune system, in agreement with the extensive literature data on their key role in envelope protein functionality [31,67]. Therefore, the correlation between sequence length variability and number of PNGSs is likely to affect the population of shifting PNGS. Zhang et al. have described these PNGS as being positioned in hot spot regions [31]. In agreement with these findings, our data show an extensive presence of shifting PNGSs in the V1 and V4, and, to a less extent, in the V2 loops that accumulate in the chronic disease stage and cluster in the C-terminal portion of the domains. These results are also in line with previously published studies on subtype A, B and C viruses, which describe an increase of the PNGSs number after the early-infection stage in the V1-V2 domain [38,42,43,68], although our data demonstrate that the distribution of "hot-spot" areas of shifting PNGS is different. Nevertheless, since HIV-1 subtype C is phylogenetically more related to clade A than clade B, our data support the possibility of evolution of a subtype-specific glycan shield by the shifting PNGSs within the constraints posed by glycan-dependent envelope functionality mediated by the more conserved PNGSs [27,31,46,69,70].

Taken altogether our data on sequence length and PNGS number variation suggest a major effect of the positive selective pressure driven by the host immune system on the V1 and V4 loops compared to the V2 and V5 loops, indicating that the V1 and V4 regions could be among the main drivers of clade C HIV-1 virus evolution. This hypothesis is also in line with recent studies that identify the V1 loop as the major regulator of virus sensitivity towards neutralizing antibodies in subtype B [71] and point to the V4 loop of subtype C as a being a major target of neutralization activity [29].

Furthermore, our results suggest that the disease stage-related envelope evolution trend, previously observed in intra-individual dynamics, could be a distinctive mechanism of HIV-1 subtype C virus escape at the population level in Southern Africa.

It has been reported that differences in aminoacid charges correlate with resistance to neutralising antibodies [68]. In our study the Q_{tot} of V1, V2 and V4 regions was found to be stable, throughout the course of the disease. Noteworthy, the V5 loop was characterized by a decrease of Q_{tot} in the transition from early to chronic disease stage followed by an increase in the transition from chronic to late disease stage. This was associated with converse changes in Q_{pos} , statistically significant in the transition from chronic to late disease stage. We can speculate that the V5 domain, while conserved in its sequence length and PNGSs number, undergoes significant changes in its net electrical charge, suggesting a crucial role of V5 charges in the 3D structure of Env, which could influence glycan packing and immune evasion, as it has been proposed previously [27-29].

In summary, our results have demonstrated the existence of a disease stage-related trend in the evolution of subtype C gp120 Env.

Increased understanding of the exposure of different Env regions that have key function in virus infectivity and transmission may help current efforts to develop a preventative vaccine against HIV infection.

Sequence Data

All the 72 sequences generated from this study were deposited in GenBank under accession numbers JN120974-JN121117

Acknowledgments

We thank Dr. Mario Falchi for his help in figure editing, Mrs. Guendalina Fornari Luswergh and Mrs Stefania Ceccarelli for her editorial assistance. We also thank Mrs Emanuela Salvi, Mrs. Claudia Rovetto and Mrs. Patrizia Di Zeo for their outstanding technical assistance. We are indebted to Dr. van Regenmortel for contributing to the discussion of the results of this work.

References

1. (2011) UNAIDS World AIDS day Report.
2. www.unaids.org
3. Rehle TM, Hallett TB, Shisana O, Pillay-van Wyk V, Zuma K, et al. (2010) A decline in new HIV infections in South Africa: estimating HIV incidence from three national HIV surveys in 2002, 2005 and 2008. *PLoS One* 5: e11094.
4. Swaziland Demographic and Health Survey 2006-2007. Central Statistical Office, Ministry of Economic Planning and Development, Swaziland.
5. <http://www.avert.org/aroundworld.htm>
6. Gonzalez N, Alvarez A, Alcami J (2010) Broadly neutralizing antibodies and their significance for HIV-1 vaccines. *Curr HIV Res* 8: 602-612.
7. Korber B, Gnanakaran S (2009) The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. *Curr Opin HIV AIDS* 4: 408-417.
8. Leitner, T, Foley B, Hahn B, Marx P, McCutchan F, et al. (2005) HIV Sequence Compendium. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, vol. LA-UR 06-0680.
9. Zhang H, Hoffmann F, He J, He X, Kankasa C, et al. (2005) Evolution of subtype C HIV-1 Env in a slowly progressing Zambian infant. *Retrovirology* 2: 67.
10. Arien KK, Vantham G, Arts EJ (2007) Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* 5: 141-151.
11. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, et al. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303: 2019-2022.
12. Sattentau QJ (1996) The role of the envelope glycoprotein in HIV transmission and pathogenesis. *PERSPECTIVE IN DRUG DISCOVERY AND DESIGN* 5: 1-16.
13. Cheng-Mayer C, Brown A, Harouse J, Luciw PA, Mayer AJ (1999) Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant in vivo by virtue of sequence changes in the extracellular envelope glycoprotein that modify N-linked glycosylation. *J Virol* 73: 5294-5300.
14. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393: 648-659.
15. Diskin R, Marcovecchio PM, Bjorkman PJ (2010) Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nat Struct Mol Biol* 17: 608-613.
16. Borggren M, Repits J, Kuylenstierna C, Sterjovski J, Churchill MJ, et al. (2008) Evolution of DC-SIGN use revealed by fitness studies of R5 HIV-1 variants emerging during AIDS progression. *Retrovirology* 5: 28.
17. Koito A, Harrowe G, Levy JA, Cheng-Mayer C (1994) Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. *J Virol* 68: 2253-2259.
18. Krachmarov C, Pinter A, Honnen WJ, Gorny MK, Nyambi PN, et al. (2005) Antibodies that are cross-reactive for human immunodeficiency virus type 1 clade a and clade B v3 domains are common in patient sera from Cameroon, but their neutralization activity is usually restricted by epitope masking. *J Virol* 79: 780-790.

19. Krachmarov CP, Honnen WJ, Kayman SC, Gorny MK, Zolla-Pazner S, et al. (2006) Factors determining the breadth and potency of neutralization by V3-specific human monoclonal antibodies derived from subjects infected with clade A or clade B strains of human immunodeficiency virus type 1. *J Virol* 80: 7127-7135.
20. Pinter A, Honnen WJ, He Y, Gorny MK, Zolla-Pazner S, et al. (2004) The V1/V2 domain of gp120 is a global regulator of the sensitivity of primary human immunodeficiency virus type 1 isolates to neutralization by antibodies commonly induced upon infection. *J Virol* 78: 5205-5215.
21. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, et al. (2003) The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol* 77: 4449-4456.
22. Doms RW, Peiper SC (1997) Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. *Virology* 235: 179-190.
23. Gnanakaran S, Lang D, Daniels M, Bhattacharya T, Derdeyn CA, et al. (2007) Clade-specific differences between human immunodeficiency virus type 1 clades B and C: diversity and correlations in C3-V4 regions of gp120. *J Virol* 81: 4886-4891.
24. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, et al. (2002) Diversity considerations in HIV-1 vaccine selection. *Science* 296: 2354-2360.
25. Cecilia D, Kulkarni SS, Tripathy SP, Gangakhedkar RR, Paranjape RS, et al. (2000) Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virology* 271: 253-258.
26. Tscherning C, Alaeus A, Fredriksson R, Björndal A, Deng H, et al. (1998) Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* 241: 181-188.
27. Wei X, Decker JM, Wang S, Hui H, Kappes JC, et al. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422: 307-312.
28. Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, et al. (2005) Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci U S A* 102: 18514-18519.
29. Moore PL, Gray ES, Choge IA, Ranchohe N, Mlisana K, et al. (2008) The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *J Virol* 82: 1860-1869.
30. Myers G, MaInnes K, Korber B (1992) The emergence of simian/human immunodeficiency viruses. *AIDS Res Hum Retroviruses* 8: 373-386.
31. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, et al. (2004) Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14: 1229-1246.
32. Polonoff E, Machida CA, Kabat D (1982) Glycosylation and intracellular transport of membrane glycoproteins encoded by murine leukemia viruses. Inhibition by aminoacid analogues and by tunicamycin. *J Biol Chem* 257: 14023-14028.
33. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A (1997) The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. *J Cell Biol* 139: 613-623.
34. Moore PL, Ranchohe N, Lambson BE, Gray ES, Cave E, et al. (2009) Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog* 5: e1000598.
35. Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, et al. (2001) N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J Biol Chem* 276: 13433-13441.
36. Lin G, Simmons G, Pohlmann S, Baribaud F, Ni H, et al. (2003) Differential N-linked glycosylation of human immunodeficiency virus and Ebola virus envelope glycoproteins modulates interactions with DC-SIGN and DC-SIGNR. *J Virol* 77: 1337-1346.
37. Botarelli P, Houlden BA, Haigwood NL, Servis C, Montagna D, et al. (1991) N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. *J Immunol* 147: 3128-3132.
38. Sagar M, Wu X, Lee S, Overbaugh J (2006) Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. *J Virol* 80: 9586-9598.
39. Sagar M, Kirkegaard E, Lavreys L, Overbaugh J (2004) Primary HIV-1 Infection is Clonal in a Minority of Women from Africa. In Conference on Retroviruses and Opportunistic Infections 11th: Program Abstr Conf Retrovir Oppor Infect 11th 2004 San Franc Calif. 11: abstract no. 385.
40. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, et al. (2009) Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol* 83: 3556-3567.
41. Coetzer M, Cilliers T, Papathanasopoulos M, Ramjee G, Karim SA, et al. (2007) Longitudinal analysis of HIV type 1 subtype C envelope sequences from South Africa. *AIDS Res Hum Retroviruses* 23: 316-321.
42. Rong R, Bibollet-Ruche F, Mulenga J, Allen S, Blackwell JL, et al. (2007) Role of V1V2 and other human immunodeficiency virus type 1 envelope domains in resistance to autologous neutralization during clade C infection. *J Virol* 81: 1350-1359.
43. Bunnik EM, Pisas L, van Nunen AC, Schuitemaker H (2008) Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J Virol* 82: 7932-7941.
44. Wyatt R, Moore J, Accola M, Desjardin E, Robinson J, et al. (1995) Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J Virol* 69: 5723-5733.
45. Peut V, Kent JS (2006) Fitness constraints on immune escape from HIV: Implications of envelope as a target for both HIV-specific T cells and antibody. *Curr HIV Res* 4: 191-197.
46. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, et al. (2002) HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* 420: 678-682.
47. Naganawa S, Yokoyama M, Shiino T, Suzuki T, Ishigatsubo Y, et al. (2008) Net positive charge of HIV-1 CRF01_AE V3 sequence regulates viral sensitivity to humoral immunity. *PLoS One* 3: e3206.
48. Suligoi B, Buttò S, Galli C, Bernasconi D, Salata RA, et al. (2008) Detection of recent HIV infections in African individuals infected by HIV-1 non-B subtypes using HIV antibody avidity. *J Clin Virol* 41: 288-292.
49. Bernasconi D, Tivoschi L, Regine V, Raimondo M, Gama D, et al. (2010) Identification of recent HIV infections and of factors associated with virus acquisition among pregnant women in 2004 and 2006 in Swaziland. *J Clin Virol* 48: 180-183.
50. Sanger F, Nicklen S, Coulson AR (1992) DNA sequencing with chain terminator inhibitors. 1977. *Biotechnology* 24: 104-108.
51. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
52. Hall TA (1999) Bioedit a user-friendly biological sequence alignment editor and analysis program for Windows 95/98 NT. *Nucleic Acids Symp Ser* 41: 95-98.
53. <http://paup.csit.fsu.edu/downl.html>
54. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
55. Jukes T, Cantor CR (1969) Evolution of protein molecules. In *Mammalian protein metabolism*. Edited by Munro HN. New York: Academic Press; 21-132.
56. Bhatta H, Goldys EM (2009) Quantitative characterization of different strains of *Saccharomyces* yeast by analysis of fluorescence microscopy images of cell populations. *J Microbiol Methods* 77: 77-84.
57. Papathanasopoulos MA, Vardas E, Wallis C, Glashoff R, Buttò S, et al. (2010) Characterization of HIV type 1 genetic diversity among South African participants enrolled in the AIDS Vaccine Integrated Project (AVIP) study. *AIDS Res Hum Retroviruses* 26: 705-709.
58. Dehò L, Walwema R, Cappelletti A, Sukati H, Sibandze D, et al. (2008) Subtype assignment and phylogenetic analysis of HIV type 1 strains in patients from Swaziland. *AIDS Res Hum Retroviruses* 24: 323-325.
59. Jacobs GB, Loxton AG, Laten A, Robson B, van Rensburg EJ, et al. (2009) Emergence and diversity of different HIV-1 subtypes in South Africa, 2000-2001. *J Med Virol* 81: 1852-1859.
60. Patel MB, Hoffman NG, Swanstrom R (2008) Subtype-specific conformational

- differences within the V3 region of subtype B and subtype C human immunodeficiency virus type 1 Env proteins. *J Virol* 82: 903-916.
61. Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, et al. (2006) Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. *J Virol* 80: 5211-5218.
62. Rong R, Li B, Lynch RM, Haaland RE, Murphy MK, et al. (2009) Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. *PLoS Pathog* 5: e1000594.
63. Kitrinou KM, Hoffman NG, Nelson JA, Swanstrom R (2003) Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection. *J Virol* 77: 6811-6822.
64. Moore PL, Gray ES, Sheward D, Madiga M, Ranchobe N, et al. (2011) Potent and Broad Neutralization of HIV-1 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Residues in the V2 Loop. *J Virol* 85: 3128-3141.
65. Riddle TM, Shire NJ, Sherman MS, Franco KF, Sheppard HW, et al. (2006) Sequential turnover of human immunodeficiency virus type 1 env throughout the course of infection. *J Virol* 80: 10591-10599.
66. Chakabva MS Patterns of N-linked glycosylation site variation in the HIV Envelope Glycoprotein Current AIDS day: challenging the crisis. Harvard School of some PNGS Public Health AIDS Initiative.
67. Go EP, Chang Q, Liao HX, Sutherland LL, Alam SM, et al. (2009) Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J Proteome Res* 8: 4231-4242.
68. Blay WM, Gnanakaran S, Foley B, Doria-Rose NA, Korber BT, et al. (2006) Consistent patterns of change during the divergence of human immunodeficiency virus type 1 envelope from that of the inoculated virus in simian/human immunodeficiency virus-infected macaques. *J Virol* 80: 999-1014.
69. Li Y, Luo L, Rasool N, Kang CY (1993) Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. *J Virol* 67: 584-588.
70. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, et al. (2005) Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol* 79: 6528-6531.
71. Ching LK, Vlachogiannis G, Bosch KA, Stamatatos L (2008) The first hypervariable region of the gp120 Env glycoprotein defines the neutralizing susceptibility of heterologous human immunodeficiency virus type 1 isolates to neutralizing antibodies elicited by the SF162gp140 immunogen. *J Virol* 82: 949-956.

This article was originally published in a special issue, **Vaccine research: HIV** handled by Editor(s). Dr. Marc Van Regenmortel, University of Strasbourg, France