Published in final edited form as:

J Acquir Immune Defic Syndr. 2015 December 15; 70(5): 463–471. doi:10.1097/QAI.0000000000000802.

Worldwide genetic features of HIV-1 Env $\alpha 4\beta 7$ binding motif: the Local Dissemination Impact of the LDI tripeptide

Sabrina H. Hait^{1,2}, Esmeralda A. Soares², Eduardo Sprinz³, James Arthos⁴, Elizabeth S. Machado^{1,5}, and Marcelo A. Soares^{1,2,*}

¹Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

²Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, Brazil

³Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Brazil

⁴Laboratory of Immune Regulation, National Institutes of Health, Bethesda, U.S.A

⁵Instituto de Puericultura e Pediatria Martagão Gesteira, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Background—HIV-1 gp120 binds to integrin $\alpha 4\beta 7$, a homing receptor of lymphocytes to gut-associated lymphoid tissues. This interaction is mediated by the LDI/V tripeptide encoded in the V2-loop. This tripeptide mimics similar motifs in MAdCAM and VCAM, the natural ligands of $\alpha 4\beta 7$. In this study we explored the association of V2-loop LDI/V mimotopes with transmission routes and patterns of disease progression in HIV-infected adult and pediatric patients. HIV-1 env sequences available in the Los Alamos HIV Sequence Database were included in the analyses.

Methodology—HIV-1 V2-loop sequences generated from infected adults and infants from South and Southeast Brazil, and also retrieved from the Los Alamos Database, were assessed for $\alpha 4\beta 7$ binding tripeptide composition. Chi-Square/Fisher's Exact test and Mann Whitney U-test were used for tripeptide comparisons. Shannon entropy was assessed for conservancy of the $\alpha 4\beta 7$ tripeptide mimotope.

Results—We observed no association between the tripeptide composition or conservation and virus transmission route or disease progression. However, LDI was linked to successful epidemic dissemination of HIV-1 subtype C in South America, and further to other expanding non-B subtypes in Europe and Asia. In Africa, subtypes showing increased LDV prevalence evidenced an ongoing process of selection towards LDI expansion, an observation also extended to subtype B in the Americas and Western Europe.

Conclusions—The V2-loop LDI mimotope was conserved in HIV-1C from South America and other expanding subtypes across the globe, which suggests that LDI may promote successful dissemination of HIV at local geographic levels by means of increased transmission fitness.

*Corresponding author: Programa de Oncovirologia, Instituto Nacional de Câncer, Rua André Cavalcante, 37 / 4º andar, Bairro de Fátima, 20231-050 Rio de Janeiro, RJ, Brazil, Tel: 55-21-3207-6591, masoares@inca.gov.br.

Conflicts of Interest: For the remaining authors none were declared.

Keywords

HIV-1; LDI tripeptide; $\alpha 4\beta 7$; virus dissemination; pathogenesis

INTRODUCTION

Integrins comprise a superfamily of twenty-four heterodimeric transmembrane proteins which are essential for cell homing to and retention within specific sites, contributing significantly to the architecture of most tissues.^{1,2} Integrin heterodimers are formed by the specific association of one of 18 α and one of eight β subunits.³ Stimulation of lymphocytes followed by their migration to lymphoid tissues, a mechanism required for most immune responses, relies on integrin activation and recognition.⁴

The $\alpha 4\beta 7$ integrin ($\alpha 4\beta 7$) plays a pivotal role in lymphocyte homing and retention to the gut-associated lymphoid tissue (GALT). Homing is mediated by VCAM and MAdCAM-1, the natural ligands of $\alpha 4\beta 7$, which share a related tripeptide in which a central aspartic acid (Asp) residue is invariable. In both cases, the Asp is flanked by an aliphatic residue; in VCAM the tripeptide appears as IDS,⁵ while in MAdCAM it appears as LDT.⁶ The LDI/V motif is not a direct contact site with its natural ligands. Rather, the Asp in each of these three ligands mediates metal cation coordination between the ligand and $\alpha 4\beta 7$. HIV-1, which replicates at high levels in GALT during the acute phase of infection, encodes a similar tripeptide motif in the V2 loop of the gp120 envelope protein (HXB2 *env* gene codons 179–181),⁷ which also mediates binding to $\alpha 4\beta 7$. Although $\alpha 4\beta 7$ is not required for infection of CD4⁺ T cells,⁸ it has been suggested that the interaction between lentiviral envelope proteins and $\alpha 4\beta 7$ facilitates the migration of virus-infected $\alpha 4\beta 7^+$ /CD4⁺ T cells to GALT and may play an important role during acute infection.^{7,9}

While the core Asp at position 180 is among the most highly conserved residues in HIV-1 gp120,¹⁰ variation at position 179 and in particular at position 181 has been reported. For example, subtype C gp120s frequently encode LDI, while subtypes B and D gp120s frequently encode LDV.⁷ Recently, He *et al.*¹¹ have suggested that the conservation of the LDI/LDV tripeptide may influence the epidemic spread of distinct HIV-1 genetic forms in China. It should be noted however that it is not known whether and how the variation at position 181 impacts gp120 binding to $\alpha 4\beta 7$. Position 181 is noteworthy in one other respect. In the RV144 vaccine trial, which showed a protective efficacy of 31.2 %, a sieve analysis of vaccinees who became infected revealed that viruses encoding position 181 residues other than the isoleucine were preferentially transmitted and thus escaped protection responses elicited by the vaccine.^{12,13} Variation at position 181 is thus worth further characterization.

HIV-1 genetic forms are heterogeneously disseminated worldwide. Nine pure genetic subtypes and over 70 circulating recombinant forms (CRFs) have been identified to date.¹⁴ Despite such diversity, specific HIV-1 subtypes/CRFs are successfully disseminated in distinct geographic areas relative to others. In South America (SA), HIV-1 subtype B predominates in the AIDS epidemic, but other subtypes such as F1 and C, and BF1 and BC recombinant forms are on the rise.^{15–18} In the present study, we have assessed the HIV-1 *env*

V2 $\alpha 4\beta 7$ binding LDI/V tripeptide composition of viruses circulating in south and southeastern Brazil (where over 80% of HIV infections in the country are located). We provide evidence that the LDI tripeptide is expanding across the continent, suggesting a preference for the transmission of viruses encoding an isoleucine at position 181. Analysis of HIV-1 *env* V2 sequences from other parts of the world retrieved from the Los Alamos Database further corroborates the hypothesis of local hegemony of distinct HIV-1 subtypes carrying the LDI variant.

MATERIALS AND METHODS

Patients and Samples

Samples from 27 HIV-infected infants followed-up since birth at the Pediatric Unit of Universidade Federal do Rio de Janeiro, RJ (southeast Brazil) and from 68 HIV-infected adults followed-up at the HIV/AIDS outpatient clinic of Hospital de Clínicas de Porto Alegre, RS (southern Brazil) were included in this study. Clinical and epidemiological data from patients of both cohorts were available and patients were classified according to likely infection route and disease progression profiles (Supplementary Tables 1 and 2). Vertically-infected children (VIC) were part of a previous study and disease progression patterns were defined using HIV viral load measurements and CD4⁺ T-cell counts according to established criteria.^{19–21} Among VIC, six samples were identified as long-term non-progressors (LTNP), with 8 years of infection and presenting CD4⁺ T-cell counts $\geq 25\%$ without ARV therapy, whereas the remaining were classified as progressors. Among infected adults (IA), 18 subjects have been identified as intravenous drug users (IDU), thus categorized as parenterally-exposed/infected. In this group, nine individuals were classified as LTNP or did not progress to AIDS before 8 years of follow-up (probable LTNP) according to established definitions.²² IAs naïve to treatment at sampling time with less than 10 years of follow-up were categorized as undetermined progressors and were excluded from progression analyses. All patients or child guardians have signed a consent form and agreed to participate to this research, and the study has been approved by the Institutional Review Boards of both University centers involved. The use of previously collected infant samples has been granted by the IRB of UFRJ.

PCR, DNA sequencing and sequence analysis

Blood samples had their peripheral blood mononuclear cells (PBMC) isolated using Ficoll density gradient. PBMC were used for genomic DNA extraction using the QIAGEN *Blood Mini Kit* (QIAGEN, Chatsworth, CA) according to the manufacturer's specifications. HIV proviral sequences comprising the HIV-1 gp120 V1–V2 hypervariable regions were PCR-amplified using a semi-nested protocol as described elsewhere.²³ Cycling conditions were the same for the first and second rounds, except for varying the annealing temperature according to the primers used (Supplementary Table 3). DNA sequencing was carried out in an automated ABI 3130XL Genetic Analyzer (Life Technologies) and manually edited with SeqMan v.7.0 (DNASTAR Inc, Madison, U.S.A.). Sequences were then aligned using BioEdit v.7.0.²⁴ Deduced Env amino acid sequences were compared regarding positions 179–181 in HXB2, corresponding to the putative integrin $\alpha 4\beta 7$ binding site at the V2 loop.

Characterization of HIV-1 recombinant forms in samples obtained in this study was limited, since sequencing was performed only for the V1–V2 loops. Therefore, sequences were classified only according to the subtype of V1–V2 hypervariable loop using the HIV Blast online tool (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). For subtype confirmation, sequences were codon-aligned with HIV-1 subtypes B, C and F reference sequences retrieved from the Los Alamos Database. A Neighbor-Joining tree was constructed using MEGA v.5.0²⁵ and the Kimura-2-parameter model was used. Recombinants remained classified as BC or BF when Blast and phylogenies could not show accurate resolution for V2 subtype and sequences were highly similar to the recombinant forms as suggested by the HIV Blast searches (identity ~80%). HIV-1 subtype definition for the retrieved sequences was the one provided at the database website.

For depicting temporal trends in the prevalence of LDI and LDV-containing HIV-1 sequences, these latter were stratified into 5-year groups (1980, 1981–1985, 1986–1990, 1991–1995, 1996–2000, 2001–2005 and 2006–2010, or into a small number of groups, depending on sequence availability) according to the collection date depicted in the Los Alamos HIV Sequence Database.

All HIV-1 *env* V2 nucleotide sequences herein generated have been deposited at the GenBank database and have been assigned the accession numbers KM102250 to KM102339.

Statistical Analyses

Shannon diversity indexes²⁶ were estimated to quantify the degree of diversity within the 179–181 tripeptide domain, using the *Shannon Entropy-One* online tool available at the Los Alamos Database website (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html). LDI/V and other variant tripeptide distribution and prevalence were compared using Chi-squared and Fisher's Exact tests in SPSS.

RESULTS

α 4 β 7 mimotope tripeptide variability among HIV forms circulating in South America

We generated V2 sequences of 23 subtype B, three subtype F1 and one BF1 recombinant virus from the VIC group, a ratio consistent with the HIV molecular epidemiology scenario in southeastern Brazil.^{27,28} Of those, 23 subjects could be classified with respect to their disease progression profiles. For the IA group, 34 subtype B and an equal number of subtype C sequences has been determined, reflecting the high occurrence of HIV-1 C in southern Brazil compared to the rest of the country.^{18,29–31} All subjects in this group were classified according to transmission route and disease progression pattern. There was no specific association between the α 4 β 7-binding LDV/I tripeptide and transmission or progression status. The LDV/I tripeptide composition was also evaluated within individual subtypes (subtype B, n = 57; C, n = 34 and F1, n = 3). Additional SA V2 sequences have been retrieved from Los Alamos (171 subtype B, 12 C, 16 F1, 130 BF and 12 BC) for analysis. All α 4 β 7 binding tripeptides experimentally obtained in our study groups harbored the aspartic acid at position 180, as well as a striking majority of SA sequences retrieved from the

Los Alamos database, regardless of the viral subtype or transmission route. However, clear differences in the variability of positions 179 and 181 were observed between SA variants. An analysis comprising all SA HIV-1 V2 sequences showed increased variation in subtype B, whilst subtypes C and F1 presented fewer non-LDI/V tripeptide variants (Table 1). BF recombinants also revealed a greater number of tripeptide variants when compared to BC (Table 1). Env amino acid positions 179–181 were assessed for conservancy on the tripeptide mimotope among HIV-1 genetic forms. We observed a higher heterogeneity at position 179 in subtypes B and BFs, whilst subtypes C, F1 and BCs presented lower variability at that position (Supplementary Figure S1A). Valine and isoleucine were the most prevalent residues observed at position 181 (59% and 35%, respectively) in subtype B, while all subtype F1 sequences presented isoleucine at position 181. Subtype C, BC and BF recombinants showed varied prevalence of Ile181 (Supplementary Figure S1B), but all higher than subtype B.

The Shannon diversity of positions 179–181 was calculated for all HIV-1 subtypes/recombinants. A higher heterogeneity degree was observed at positions 179 and 181 in subtype B and BF recombinants, whereas subtypes C and F1 showed higher conservation at those positions (Table 2).

LDI favors dissemination of HIV-1 non-subtype B forms in South America

As shown above, the proportion of LDV/I was equivalent for subtypes B, C, F1 and BC and BF recombinants, even when data retrieved from Los Alamos Databases was included. No statistical significance in LDV/I distribution was observed for any particular HIV-1 genetic form, indicating that both LDV and LDI are maintained as potential $\alpha 4\beta 7$ integrin putative binding mimotopes (Supplementary Figure S2). Notwithstanding, there was a clear difference in LDI prevalence over LDV when both tripeptides were analyzed separately. LDI variants were markedly prevalent in the V2 loop of subtypes C and F1, but not in subtype B, which showed a higher prevalence of LDV (Figure 1A). The addition of V2 sequences retrieved from the Los Alamos data base corroborated our results, evidencing a high prevalence of the LDI in subtypes C, F, BF and BC, but not in subtype B (Figure 1B). Of note, epidemiological data suggests that all those HIV-1 non-B forms are on the rise in SA (except for pure subtype F1, which is thought to be extinct on the continent), suggesting that the LDI tripeptide composition may promote HIV dissemination of those variants in the continent.

Prevalence of LDI and dissemination of HIV-1 subtypes across the globe

Our data from SA suggested that the LDI tripeptide variation could be linked to a more successful HIV epidemic dissemination. We wanted to further test this hypothesis by analyzing HIV-1 V2 tripeptide from several other parts of the world. Viral *env* V2 loop sequences across continents were retrieved from the Los Alamos Database, and we assessed conservation and prevalence of the LDI $\alpha 4\beta 7$ -binding mimotope in HIV-1 forms undergoing expansion in various locales, compared to sequences from subtypes previously established at those locations.

Based on data reported by He *et al.*¹¹ in China, we verified that LDI accounts for 94% of $\alpha 4\beta 7$ binding domains in the fastest disseminating HIV-1 CRF07_BC viruses. We also verified a high prevalence of LDI in the spreading CRF01_AE and CRF08_BC forms (69.5% and 84%, respectively; Table 3). In contrast, only 45% of subtypes B viruses carried LDI, while the majority encoded LDV or other variants.

In Southeast Asia, we observed a striking prevalence of the recombinant form CRF01_AE, consistent with recent literature.³² Of 601 HIV-1 V2 sequences originated from those countries, 465 (77%) were CRF01_AE and 77% of them presented LDI (Table 3). In contrast, LDI prevalence was only 14% among subtype B sequences, lower than that of LDV (56%).

In Western Europe, mounting evidence of the rise of non-B HIV-1 subtypes, including CRF01_AE and subtypes C, F1 and G has been published.^{33–39} The estimated distribution of LDI in these HIV-1 forms reaches 59–89%. Taking together, LDI prevails among emerging HIV subtypes in Europe in areas previously marked by subtype B epidemics.

In the former Soviet Union (FSU), introduction of HIV started during the 80's among MSM infected with subtype B, determining the hegemony of this subtype for over a decade.^{40,41} Currently, subtype A/A1 accounts for 71% of the viruses in the region and its LDI prevalence is 97%, while LDV accounts for only 0.7% (Table 3). Subtype B sequences from FSU, although very limited, showed equivalent rates of LDI and LDV (approximately 30% each).

Since sub-Saharan Africa presents an extremely diverse distribution of HIV variants,⁴² we focused our evaluation of the LDI distribution in HIV-1 forms according to their predominance in the South/East and West/Central regions. In the former, HIV-1C accounts for over 95% of infections.⁴³ In South Africa, LDI accounted for 42% of HIV-1C sequences, whereas LDV prevalence was only 6% (Table 3). In East Africa, recent surveys showed dissemination of subtype A/A1 in Kenya, Uganda and Rwanda. In Rwanda, subtype A/A1 already accounts for 95% of HIV epidemics.^{44–46} Unexpectedly, HIV-1A LDI prevalence was slightly below that of LDV (41% *versus* 48%). HIV-1C showed a striking prevalence of LDI over LDV, while HIV-1D evidenced distinctly larger distribution of LDV (78%) compared to LDI (7%) (Table 3).

In West/Central Africa, all HIV-1 forms co-circulate.⁴⁷ Despite such diversity, CRF02_AG has recently presented evident dissemination, spreading faster than other HIV variants in Cameroon, Senegal and Ghana.^{48,49} V2 sequences from CRF02_AG, which belong to subtype A, were separately analyzed in West and Central Africa. For western African countries, CRF02_AG accounted for 58%. LDI prevalence in CRF02_AG from this region was low (14%) compared to LDV (49%). In Cameroon and Central African Republic, CRF02_AG contributed to 19% of all infections, and the LDI frequency was similar to that of Western Africa. In contrast, the LDI frequency in CRF11_cpx, currently expanding in Cameroon and in the Central African Republic, revealed a higher frequency of LDI (Table 3).

LDI outgrows LDV over time in the epidemics

The low prevalence of LDI among HIV-1A/A1 sequences in East Africa and CRF02_AG sequences from West/Africa, where these respective genetic forms predominate, was unexpected. Moreover, even with the increasing prevalence of non-B subtypes in North and South America, as well as in Western Europe, HIV-1B still predominates in the epidemic, with lower LDV prevalence. One alternative explanation for those observations could be that a shift of LDI over LDV over time is still going on, while the retrieved sequences from the Los Alamos database represent only a snapshot that includes older sequences. To evaluate this scenario, we analyzed LDI/LDV prevalence over time in those geographic areas using the sample collection dates available in the database. HIV-1A/A1 sequences of East Africa, from 1990 to 2000, evidenced a higher prevalence of LDV (71%) over LDI (29%). In contrast, LDI sequences showed elevated prevalence (69%) over LDV forms in this location from 2000 to 2010 (Table 4). In West and Central Africa, although LDV maintained high prevalence over LDI HIV CRF02_AG, increased frequencies of LDI were observed after 2000 (24% and 27%, respectively) compared to sequences sampled from 1990–2000 (19 and 17%, respectively) (Supplementary Table 4). Most strikingly, when analyzing data from the Americas and Western Europe, from where a great number of sequences are available, a very clear trend for an increase in time of LDI over LDV in HIV-1B is observed (Figure 2). Taken altogether, our data suggest that although LDV is still predominant in some areas, it is being replaced by LDI over time likely due to increased transmission potential of viruses encoding the latter.

DISCUSSION

In the present study, we assessed the HIV-1 *env* V2 region of viral subtypes circulating in South America, focusing on the recently described $\alpha 4\beta 7$ integrin binding LDI/V tripeptide. We show that LDI either already predominates or is expanding over LDV. We provide evidence that similar trends may be occurring across the globe in various HIV-1 subtypes/CRFs.

It has been consistently shown that the HIV envelope undergoes different types of selective pressures at mucosal transmission *versus* during the course of infection. Mucosal transmission is inefficient, and in order to reach target cells, the virus is required to break through the physical barriers of the mucosa and enter target cells that can support productive infection. This represents the first selective pressure in sexual transmission.^{50–54} As a consequence, the productive infection that follows primary contact generally reflects an expansion from a single founder virus with a particular phenotype.⁵⁵ Although *env* selection has been largely investigated in different infection routes in an isolated manner, little is known about the way in which viral signatures of transmission are related to different routes of transmission. Nor is it understood why in many parts of the world HIV-1 subtype C appears to be replacing other subtypes. HIV-1C, which typically encodes LDI, is the prevalent subtype in the global AIDS epidemic and is increasing in South America. Moreover, it has been suggested that the interaction between gp120 and $\alpha 4\beta 7$ may facilitate mucosal transmission, which is consistent with the observation that $\alpha 4\beta 7$ *high*/CCR5*high* activated CD4⁺ T-cells are the lymphocyte subset preferentially infected by HIV.⁵⁶ We

speculate that some early transmitting viral gp120s bear structural features that exploit $\alpha 4\beta 7$ interactions allowing them to gain access to the GALT, where cells that are highly susceptible to productive infection are readily available. We wonder whether the LDI tripeptide confers greater ability to interact with the $\alpha 4\beta 7$ integrin, and this might explain the high prevalence of LDI in SA HIV-1C V2 sequences and also in other disseminating non-B HIV-1 forms in SA despite the observation that V2- $\alpha 4\beta 7$ interaction does not appear essential for viral infectivity *per se*.⁵⁷

To further test our hypothesis of an increased epidemic potential provided by the LDI tripeptide observed in SA, we expanded our analysis to gp120 sequences from other areas of the world. We found evidence suggesting that HIV forms carrying the LDI tripeptide variant are expanding elsewhere. For instance, CRF01_AE, which originated in Central Africa and predominantly encodes LDI, has disseminated to countries such as Thailand, Vietnam, Malaysia, Cambodia, Indonesia and Singapore, where previously subtype B, which typically encodes LDV, predominated in the early years of the AIDS epidemic.^{58,59} Similarly, subtype B predominated in Western Europe,^{38,60,61} but in recent years, there has been a significant expansion of non-B subtypes in that area.^{38,61} We should note however that this expansion may be explained in part by the arrival of African immigrants.⁶² Recent epidemiological studies revealed CRF01_AE expansion in countries such as Cyprus, France, Italy, Greece and Spain,^{33,34} subtype F1 in Romania³⁶ and subtype G in the Iberian Peninsula,³⁴ all showing noticeable prevalence of the LDI $\alpha 4\beta 7$ binding tripeptide. Likewise, the AIDS epidemic in FSU turned to be represented by LDI prevalent subtypes A/A1, showing greater distribution than the previously dominant form CRF03_AB. The HIV epidemics in Central/East Africa reveals a more diverse scenario, with the introduction of subtypes A/A and D.⁴² While the first showed a great prevalence of the LDI V2 tripeptide, the second displays a majority of LDV-carrying strains. HIV-1D revealed to be dominant and basically restricted to Uganda and Kenya. Similar findings were observed for CRF02_AG in West/Central Africa. Interestingly, however, a temporal analysis of sequences showed that in most of those cases, LDV is being outgrown by LDI, consistent with a scenario of epidemic spread advantage conferred by the latter tripeptide. Even more strikingly, the same scenario of outgrowth of LDI over LDV across the AIDS epidemic chronology was corroborated among HIV-1B sequences for the Americas and Western Europe, from where a large number of HIV-1 sequences were available.

The idea that an enhanced binding of LDI to $\alpha 4\beta 7$ compared to LDV came from the original observation by Arthos *et al.*⁷ that HIV-1C gp120 had a better binding when compared to gp120 derived from HIV-1A and B in an *in vitro* binding assay (see Fig. 1 of that paper) and that HIV-1C had an increased prevalence of LDI when compared to other subtypes (see Suppl. Fig 8 of the paper). However, one should note that the currently available assay to measure the binding of gp120 to $\alpha 4\beta 7$ suffers from several limitations, which would make the conclusion that LDI should bind stronger to $\alpha 4\beta 7$ just too simplistic. Firstly, the assay requires the $\alpha 4\beta 7$ to be in a high state of activation through the use of Mn^{2+} , a state unlikely to be predominantly present physiologically. Secondly, HIV-1 gp120 binds to the target cell as a trimer, a condition not reproduced in the assay, which is made with monomeric proteins. Finally, more recent work by the group of Dr. Arthos has pointed out that additional, underlying sequences in gp120 V2 and in V1 influence binding to $\alpha 4\beta 7$,⁶³ and it is unlikely

that the LDI/V tripeptide is the unique determinant to this binding. Other unexplored features of the gp120/α4β7 binding, such as binding kinetics itself and the signal transduction that follows that interaction, are also not measured in this simple assay. Overall, the currently available assay may not reflect the net impact of each individual tripeptide variant in HIV transmission, and most importantly its dissemination at an epidemic level.

To further test a possible predominance of LDI among recent HIV infections, we have assessed the V2 tripeptide composition among more than 5,000 transmitted/founder viruses published recently^{64–66} (Supplementary Fig. S3). Rather than such predominance, we found compositions that resemble the HIV-1 subtypes involved (e.g., a predominance of LDI in HIV-1C, and of LDV in HIV-1B), further supporting the hypothesis that additional determinants in Env are important for HIV transmission.

Among the putative explanations for a transition from LDV to LDI is the possibility that as various regions of gp120 have evolved, immune pressure has led to this transition. Such an explanation must however take into account the fact that recent structural analyses of near-native trimmers indicate that, in the fully unliganded state, the LDV/I is not exposed⁶⁷ and thus unlikely to be subject to immune pressure. However, the envelope structure is dynamic and variable, and therefore we cannot rule this possibility out entirely. A second consideration arguing against immune pressure directed at LDV/I is the potential for autoreactivity, insofar as either motif is likely to be a structural mimic of epitopes presented by human MAdCAM, VCAM and fibronectin. In this respect, the human immune system may recognize LDV/I as self, which would lead to the deletion of immune responses targeting this motif.

In conclusion, we provide in this report evidence that the LDI epitope is expanding over time and across the globe. This is noteworthy because this tripeptide is involved in the interaction with α4β7, which may be linked to increased transmission fitness. Moreover, the LDI epitope has been associated with the molecular sieving in the RV14 trial.¹² Although much remains to be learned about the role of the LDI/LDV epitope in HIV transmission and pathogenesis, the data presented in this report further reinforces the necessity of further investigating this important HIV-1 *env* V2 tripeptide.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Source of Funding: SHH received a PhD scholarship from the Brazilian Ministry of Education (CAPES). MAS is recipient of grant no. 304416/2010-0 from the Brazilian Ministry of Science and Technology (CNPq) and grants no. E-26/103.059/2011 and E-26/112.647/2012 from the Rio de Janeiro State Science Foundation (FAPERJ). ESM is recipient of grant 304169/2013-7 from CNPq and grant no. E26/111.689/2013 from FAPERJ.

References

1. Kamata T, Puzon W, Takada Y. Identification of putative ligand-binding sites of the integrin alpha 4 beta 1 (VLA-4, CD49d/CD29). *Biochem J.* 1995; 305(Pt 3):945–951. [PubMed: 7531439]

2. Hemler ME, Huang C, Takada Y, Schwarz L, Strominger JL, Clabby ML. Characterization of the cell surface heterodimer VLA-4 and related peptides. *J Biol Chem.* 1987; 262(24):11478–11485. [PubMed: 3497921]
3. Takada Y, Ye X, Simon S. The integrins. *Genome Biol.* 2007; 8(5):215. [PubMed: 17543136]
4. Schwartz MA, Schaller MD, Ginsberg MH. Integrins: emerging paradigms of signal transduction. *Annu Rev Cell Dev Biol.* 1995; 11:549–599. [PubMed: 8689569]
5. Clements JM, Newham P, Shepherd M, et al. Identification of a key integrin-binding sequence in VCAM-1 homologous to the LDV active site in fibronectin. *J Cell Sci.* 1994; 107(Pt 8):2127–2135. [PubMed: 7527054]
6. Viney JL, Jones S, Chiu HH, et al. Mucosal addressin cell adhesion molecule-1: a structural and functional analysis demarcates the integrin binding motif. *J Immunol.* 1996; 157(6):2488–2497. [PubMed: 8805649]
7. Arthos J, Cicala C, Martinelli E, et al. HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat Immunol.* 2008; 9(3):301–309. [PubMed: 18264102]
8. Perez LG, Chen H, Liao HX, Montefiori DC. Envelope glycoprotein binding to the integrin alpha4beta7 is not a general property of most HIV-1 strains. *J Virol.* 2014; 88(18):10767–10777. [PubMed: 25008916]
9. Byrreddy SN, Kallam B, Arthos J, et al. Targeting alpha4beta7 integrin reduces mucosal transmission of simian immunodeficiency virus and protects gut-associated lymphoid tissue from infection. *Nat Med.* 2014; 20(12):1397–1400. [PubMed: 25419708]
10. Nyambi PN, Mbah HA, Burda S, et al. Conserved and exposed epitopes on intact, native, primary human immunodeficiency virus type 1 virions of group M. *J Virol.* 2000; 74(15):7096–7107. [PubMed: 10888650]
11. He X, Mokili JL, Hong K, et al. Conservancy of the alpha4beta7 integrin mimotope in the V2 domain of HIV type 1 CRF07_BC compared to subtype B' strains in China. *AIDS Res Hum Retroviruses.* 2011; 27(10):1127–1133. [PubMed: 21417760]
12. Rolland M, Edlefsen PT, Larsen BB, et al. Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature.* 2012; 490(7420):417–420. [PubMed: 22960785]
13. Edlefsen PT, Rolland M, Hertz T, et al. Comprehensive sieve analysis of breakthrough HIV-1 sequences in the RV144 vaccine efficacy trial. *PLoS Comput Biol.* 2015; 11(2):e1003973. [PubMed: 25646817]
14. HIV Sequence Database. <http://www.hiv.lanl.gov/>. Accessed March 1, 2015
15. Santos AF, Schrago CG, Martinez AM, et al. Epidemiologic and evolutionary trends of HIV-1 CRF31_BC-related strains in southern Brazil. *J Acquir Immune Defic Syndr.* 2007; 45(3):328–333. [PubMed: 17496564]
16. Bello G, Aulicino PC, Ruchansky D, et al. Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. *Retrovirology.* 2010; 7:22. [PubMed: 20307282]
17. Carr JK, Avila M, Gomez Carrillo M, et al. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS.* 2001; 15(15):F41–47. [PubMed: 11600844]
18. Soares EA, Martinez AM, Souza TM, et al. HIV-1 subtype C dissemination in southern Brazil. *AIDS.* 2005; 19(Suppl 4):S81–86. [PubMed: 16249660]
19. Abrams EJ, Weedon J, Steketee RW, et al. Association of human immunodeficiency virus (HIV) load early in life with disease progression among HIV-infected infants. New York City Perinatal HIV Transmission Collaborative Study Group. *J Infect Dis.* 1998; 178(1):101–108. [PubMed: 9652428]
20. Levy JA. HIV pathogenesis: 25 years of progress and persistent challenges. *AIDS.* 2009; 23(2):147–160. [PubMed: 19098484]
21. Gray L, Newell ML, Thorne C, Peckham C, Levy J. Fluctuations in symptoms in human immunodeficiency virus-infected children: the first 10 years of life. *Pediatrics.* 2001; 108(1):116–122. [PubMed: 11433063]
22. Casado C, Colombo S, Rauch A, et al. Host and viral genetic correlates of clinical definitions of HIV-1 disease progression. *PLoS One.* 2010; 5(6):e11079. [PubMed: 20552027]

23. Sabino EC, Shpaer EG, Morgado MG, et al. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol.* 1994; 68(10):6340–6346. [PubMed: 8083973]
24. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999; 41:95–98.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28(10):2731–2739. [PubMed: 21546353]
26. Shannon CE. A mathematical theory of communication. *Bell System Technical Journal.* 1948; 27:623–656.
27. Pimentel VF, Morgado MG, Bello G, et al. Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. *AIDS Res Hum Retroviruses.* 2013; 29(12):1553–1561. [PubMed: 23987184]
28. Molina RM, Torina AG, Biffi K, et al. Prevalence of HIV-1 subtypes in Brazilian children with perinatally acquired infection. *J Int Assoc Physicians AIDS Care (Chic).* 2009; 8(2):106–112. [PubMed: 19264724]
29. Graf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology.* 2013; 435(1):170–178. [PubMed: 22999094]
30. Silveira J, Santos AF, Martinez AM, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J Clin Virol.* 2012; 54(1):36–41. [PubMed: 22326760]
31. Bello G, Zanotto PM, Iamarino A, et al. Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. *PLoS One.* 2012; 7(4):e35649. [PubMed: 22530062]
32. Ng KT, Ong LY, Lim SH, Takebe Y, Kamarulzaman A, Tee KK. Evolutionary history of HIV-1 subtype B and CRF01_AE transmission clusters among men who have sex with men (MSM) in Kuala Lumpur, Malaysia. *PLoS One.* 2013; 8(6):e67286. [PubMed: 23840653]
33. Brand D, Cazein F, Lot F, et al. Continuous spread of HIV-1 subtypes D and CRF01_AE in France from 2003 to 2009. *J Clin Microbiol.* 2012; 50(7):2484–2488. [PubMed: 22553246]
34. Thomson MM, Najera R. Increasing HIV-1 genetic diversity in Europe. *J Infect Dis.* 2007; 196(8):1120–1124. [PubMed: 17955428]
35. Yirrell DL, Goldberg DJ, Whitelaw J, McSharry C, Raeside F, Codere G. Viral subtype and heterosexual acquisition of HIV infections diagnosed in Scotland. *Sex Transm Infect.* 1999; 75(6):392–395. [PubMed: 10754942]
36. Paraschiv S, Foley B, Otelea D. Diversity of HIV-1 subtype C strains isolated in Romania. *Infect Genet Evol.* 2011; 11(2):270–275. [PubMed: 20620240]
37. Manolescu LS, Temereanca A, Ruta S. HIV-1 circulating subtypes in Romania. *Roum Arch Microbiol Immunol.* 2013; 72(2):121–134. [PubMed: 24187810]
38. Esteves A, Parreira R, Venenno T, et al. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Res Hum Retroviruses.* 2002; 18(5):313–325. [PubMed: 11897032]
39. Delgado E, Thomson MM, Villahermosa ML, et al. Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J Acquir Immune Defic Syndr.* 2002; 29(5):536–543. [PubMed: 11981372]
40. Bobkov A, Cheingsong-Popov R, Garaev M, et al. Identification of an env G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. *AIDS.* 1994; 8(12):1649–1655. [PubMed: 7888112]
41. Liitsola K, Laukkanen T, Denisova A, et al. Genetic characterization of HIV-1 strains in the Baltic countries and Russia. *Scand J Infect Dis.* 1996; 28(6):537–541. [PubMed: 9060052]
42. Lihana RW, Ssemwanga D, Abimiku A, Ndembi N. Update on HIV-1 diversity in Africa: a decade in review. *AIDS Rev.* 14(2):83–100. [PubMed: 22627605]
43. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS.* 25(5):679–689. [PubMed: 21297424]
44. Kemal KS, Anastos K, Weiser B, Ramirez CM, Shi Q, Burger H. Molecular epidemiology of HIV type 1 subtypes in Rwanda. *AIDS Res Hum Retroviruses.* 2013; 29(6):957–962. [PubMed: 23458210]

45. Khoja S, Ojwang P, Khan S, Okinda N, Harania R, Ali S. Genetic analysis of HIV-1 subtypes in Nairobi, Kenya. *PLoS One*. 2008; 3(9):e3191. [PubMed: 18784834]
46. Rusine J, Jurriaans S, van de Wijgert J, et al. Molecular and phylogeographic analysis of human immuno-deficiency virus type 1 strains infecting treatment-naïve patients from Kigali, Rwanda. *PLoS One*. 2012; 7(8):e42557. [PubMed: 22905148]
47. Peeters M, Toure-Kane C, Nkengasong JN. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS*. 2003; 17(18):2547–2560. [PubMed: 14685049]
48. Leye N, Vidal N, Ndiaye O, et al. High frequency of HIV-1 infections with multiple HIV-1 strains in men having sex with men (MSM) in Senegal. *Infect Genet Evol*. 2013; 20:206–214. [PubMed: 24035811]
49. Fischetti L, Opare-Sem O, Candotti D, Sarkodie F, Lee H, Allain JP. Molecular epidemiology of HIV in Ghana: dominance of CRF02_AG. *J Med Virol*. 2004; 73(2):158–166. [PubMed: 15122787]
50. Rong R, Gnanakaran S, Decker JM, et al. Unique mutational patterns in the envelope alpha 2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. *J Virol*. 2007; 81(11):5658–5668. [PubMed: 17360739]
51. Chohan B, Lang D, Sagar M, et al. 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*. 2005; 79(10):6528–6531. [PubMed: 15858037]
52. Sagar M, Laeyendecker O, Lee S, et al. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J Infect Dis*. 2009; 199(4):580–589. [PubMed: 19143562]
53. Sagar M. HIV-1 transmission biology: selection and characteristics of infecting viruses. *J Infect Dis*. 2010; 202(Suppl 2):S289–296. [PubMed: 20846035]
54. Cicala C, Arthos J, Fauci AS. HIV-1 envelope, integrins and co-receptor use in mucosal transmission of HIV. *J Transl Med*. 2011; 9(Suppl 1):S2. [PubMed: 21284901]
55. Brenchley JM, Douek DC. The mucosal barrier and immune activation in HIV pathogenesis. *Curr Opin HIV AIDS*. 2008; 3(3):356–361. [PubMed: 19372990]
56. Cicala C, Martinelli E, McNally JP, et al. The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proc Natl Acad Sci U S A*. 2009; 106(49):20877–20882. [PubMed: 19933330]
57. Cao J, Sullivan N, Desjardin E, et al. Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol*. 1997; 71(12):9808–9812. [PubMed: 9371651]
58. Perez-Losada M, Jobes DV, Sinangil F, et al. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. *PLoS One*. 2011; 6(3):e16902. [PubMed: 21423744]
59. Liao H, Tee KK, Hase S, et al. Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. *Virology*. 2009; 391(1):51–56. [PubMed: 19540543]
60. Brunet JB, Bouvet E, Massari V. Epidemiological aspects of acquired immune deficiency syndrome in France. *Ann N Y Acad Sci*. 1984; 437:334–339. [PubMed: 6598308]
61. Giuliani M, Santoro MM, Lo Presti A, et al. Circulation of HIV-1 CRF02_AG among MSM population in central Italy: a molecular epidemiology-based study. *Biomed Res Int*. 2013; 2013:810617. [PubMed: 24369538]
62. Vittecoq D, May T, Roue RT, et al. Acquired immunodeficiency syndrome after travelling in Africa: an epidemiological study in seventeen Caucasian patients. *Lancet*. 1987; 1(8533):612–615. [PubMed: 2881142]
63. Nawaz F, Cicala C, Van Ryk D, et al. The genotype of early-transmitting HIV gp120s promotes alpha (4) beta(7)-reactivity, revealing alpha (4) beta(7) +/CD4+ T cells as key targets in mucosal transmission. *PLoS Pathog*. 2011; 7(2):e1001301. [PubMed: 21383973]

64. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A*. 2008; 105(21):7552–7557. [PubMed: 18490657]
65. Sanchez AM, DeMarco CT, Hora B, et al. Development of a contemporary globally diverse HIV viral panel by the EQAPOL program. *J Immunol Methods*. 2014; 409:117–130. [PubMed: 24447533]
66. Sterrett S, Learn GH, Edlefsen PT, et al. Low Multiplicity of HIV-1 Infection and No Vaccine Enhancement in VAX003 Injection Drug Users. *Open Forum Infect Dis*. 2014; 1(2):ofu056. [PubMed: 25734126]
67. Pan R, Gorny MK, Zolla-Pazner S, Kong XP. The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel. *J Virol*. 2015 [EPub ahead of print].

Figure 1.

$\alpha 4\beta 7$ mimotope tripeptide distribution among HIV-1 V2 sequences from South America. (A) HIV-1 subtypes B and C in pediatric and adult samples (study groups). Because the number of V2 F1 subtypes was limited in our dataset ($n = 3$), it is not shown. (B) HIV-1 subtypes B, C, F1, BC and BF from South America (including sequences retrieved from the Los Alamos HIV Sequence Database).

Figure 2.

Line trends of LDI increase and LDV decrease over time in the HIV/AIDS epidemics among HIV-1 subtype B sequences retrieved from the Los Alamos HIV Sequence Database isolated and characterized in North America (A), South America (B) and Western Europe (C). The asterisks in all three panels refer to merging of periods 2006-2010 and >2010, due to low numbers of sequences deposited in the Los Alamos Database after 2010.

Table 1

α4β7 binding motif variants found in South American HIV-1 subtypes B, C and F1 and BF and BC recombinants.

Tripeptide variants	Subtype B	%	Tripeptide variants	Subtype C	%	Tripeptide variants	Subtype F1	%	Tripeptide variants	BC	%	Tripeptide variants	BF	%
LDV	122	53.5	LDI	36	78.2	LDI	16	84.1	LDI	7	58.4	LDI	75	57.8
LDI	69	30.3	LDV	4	8.7	PDI	1	5.3	LDV	3	25	LDV	32	24.6
LDL	10	4.4	undefined	2	4.3	VDI	1	5.3	IDV	1	8.3	LDL	7	5.4
SDV	3	1.3	IDI	1	2.2	LGI	1	5.3	PDI	1	8.3	IDI	2	1.5
TDV	3	1.3	VDI	1	2.2	Others*	0	0	Others*	0	0	IDM	2	1.5
IDI	3	1.3	LDM	1	2.2							Others*	12	9.2
Others*	18	7.9	TDV	1	2.2									

* Others include all variants with a frequency below 1% for a given subtype/recombinant

Table 2

Shannon entropy indexes of HIV-1 subtypes and recombinant forms circulating in South America. Data includes sequences from our study groups and from the Los Alamos HIV Sequence Database. Diversity was significantly higher in HIV-1B and BF recombinant forms for positions 179 and 181 (indicated in bold). Lower variability was observed in position 180, explained by the aspartic core maintenance.

Subtype/Recomb.Form	Shannon Entropy Index		
	<i>AA 179</i>	<i>AA 180</i>	<i>AA 181</i>
B	0.595	0.056	0.908
C	0.313	0.000	0.619
F1	0.409	0.206	0.000
BC	0.566	0.000	0.637
BF	0.595	0.000	0.939

Table 3

α 4 β 7 tripeptide mimotope distribution among worldwide HIV-1 V2 sequences retrieved from the Los Alamos HIV Sequence Database

SUBTYPE	TRIPLEPTIDE	TOTAL SEQS	%
FORMER USSR			
A/A1 (145) *	LDI	141	97.2
	LDV	1	0.7
	others	3	2.1
B (24)	LDI	7	29.2
	LDV	7	29.2
	others	10	41.7
SOUTHEAST ASIA			
CRF01_AE (465)	LDI	357	76.8
	LDV	8	1.7
	others	100	21.5
B (43)	LDI	6	14
	LDV	24	55.8
	others	13	30.2
CHINA			
CRF01_AE (282)	LDI	196	69.5
	LDV	8	2.8
	others	78	27.7
CRF07_BC (341)	LDI	321	94.1
	LDV	13	3.8
	others	7	2.1
CRF08_BC (62)	LDI	52	83.9
	LDV	1	1.6
	others	9	14.5
B (229)	LDI	103	45
	LDV	79	34.4
	others	47	20.5
SOUTH AFRICA			
C (744)	LDI	311	41.8
	LDV	48	6.4
	others	385	51.8
A/A1 (4)	LDI	0	0
	LDV	3	75
	others	1	25
B (13)	LDI	3	23.1
	LDV	8	61.5
	others	2	15.4

SUBTYPE	TRYPEPTIDE	TOTAL SEQS	%
WESTERN AFRICA			
CRF02_AG (80)	LDI	11	13.8
	LDV	39	48.8
	others	30	37.5
EAST AFRICA			
A/A1 (303)	LDI	125	41.2
	LDV	146	48.2
	others	32	10.6
C (93)	LDI	62	66.7
	LDV	15	16.1
	others	16	17.2
D (40)	LDI	3	7.5
	LDV	31	77.5
	others	6	15
CENTRAL AFRICA			
CRF02_AG (56)	LDI	7	12.5
	LDV	21	37.5
	others	28	50
CRF11_cpx (30)	LDI	13	43.3
	LDV	14	46.7
	others	3	10

* Numbers in parentheses refer to the number of sequences analyzed for each given HIV subtype and location.