AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 35, Number 6, 2019
© Mary Ann Liebert, Inc.
DOI: 10.1089/aid.2018.0267

Emerging Patterns in HIV-1 gp120 Variable Domains in Anatomical Tissues in the Absence of a Plasma Viral Load

Susanna L. Lamers,[1] Gary B. Fogel,[2] David J. Nolan,[1] Andrew E. Barbier,[1] Rebecca Rose,[1] Elyse J. Singer,[3,4] Maria Paz Gonzalez-Perez,[5] and Michael S. McGrath[6]

Abstract

The HIV envelope protein contains five hypervariable domains (V1–V5) that are fundamental for cell entry. We contrasted modifications in the variable domains derived from a panel of 24 tissues from 7 subjects with no measurable plasma viral load (NPVL) to variable domains from 76 tissues from 15 subjects who had a detectable plasma viral load (PVL) at death. NPVL subject's V1 and V2 domains were usually highly length variable, whereas length variation in PVL sequences was more conserved. Longer V1s contained more charged residues, whereas longer V2s were more glycosylated. Structural analysis demonstrated V1/V2 charge, and N-site additions/subtractions were localized to the CD4 binding pocket. Diversified envelopes in tissues during therapy may represent a mechanism for HIV persistence in tissues, as binding pocket complexity is associated with HIV that may escape neutralization, whereas shorter envelopes are associated with increased infectivity. Further analysis of tissue-derived envelope sequences may enable better understanding of potential immunological approaches targeting the persistent HIV reservoir.

Keywords: HIV envelope, tissue-based reservoirs, envelope diversity

HIV undergoes extensive diversification within an individual over the course of the infection. The HIV envelope (Env) protein, which contains two subunits gp120 and gp41, exists as a trimer on the surface of the virus. This trimeric structure serves as the point of contact between the virus and the cellular targets of infection: T-cells, macrophages (and their precursors), and dendritic cells.

As a major factor in cell entry, Env is the focus of multiple strategies for drug/vaccine development. The gp120 subunit contains five variable domains (V1–V5), of which three are remarkably elastic (V1, V2, and V4), displaying many nucleotide insertions and deletions (indels) in both intra- and intersubject sequence populations. These indels often result in the gain or loss of potential N-linked glycosylation sites (N-sites). The network of N-sites on gp120 serves as targets for broadly neutralizing antibodies[1] and contributes to immune evasion and escape during HIV-1 infection.[2]

Understanding the specific selective pressures on the virus in tissue compartments is of critical importance, as increasing evidence suggests that tissues may be a major source of viral rebound after cessation of combined antiretroviral therapy (cART). Although analysis of gp120 variable domains is hampered by extensive indels that disrupt a homologous alignment, a number of interesting observations suggest that patterns of variability reflect changing selective pressures. For example, gp120 tends to be shorter and contain fewer N-sites in early infection[3] and may be preferentially transmitted.[4] Furthermore, cloned early viruses of shorter length replicated efficiently in activated human $CD4^+$T cells but not in cultured monocyte-derived macrophages,[5] suggesting unique replication/selection pressures in tissue-based immune cells.

The flexibility of gp120 variable domains has rarely been assessed in anatomical tissues. In this study, we analyzed gp120 variable domain sequences for length, glycosylation, charge, and structure from anatomical tissues of 7 subjects who were on cART until death with no measurable plasma viral load (denoted NPVL subjects) and 15 subjects who had a

[1]BioInfoExperts LLC, Thibodaux, Louisiana.
[2]Natural Selection, Inc., San Diego, California.
[3]The National Neurological AIDS Bank, University of California, Los Angeles, California.
[4]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine and Olive View-UCLA Medical Center, Los Angeles, California.
[5]Program in Molecular Medicine, The University of Massachusetts Medical School, Worcester, Massachusetts.
[6]Departments of Laboratory Medicine, Pathology and Medicine, The University of California at San Francisco, San Francisco, California.

detectable plasma viral load at death (denoted PVL subjects). All subjects in the study were infected with HIV Subtype B.

To generate context for length variation found in our study, we downloaded and assessed the length distributions of V1 ($n = 5565$), V2 ($n = 5964$), and V4 ($n = 8828$) Subtype B sequences deposited in the Los Alamos HIV Database (www.hiv.lanl.gov) using the ''one sequence per subject'' option. While the bulk of the Los Alamos variable regions were restricted to a limited length range (V1 = 24–34 amino acids, V2 = 39–43 amino acids and V4 = 30–33 amino acids), the overall variance for V1, V2, and V4 length was considerable with lengths of 4–58, 23–75, and 6–54 amino acids for each domain, respectively, thus demonstrating the remarkable elasticity of gp120 in the overall population (Fig. 1).

NPVL and PVL subject and sequence information is provided in Table 1. All sequence data were previously published and available in GenBank (JN786871-JN786685, HM001362-HM002482, KU708874-KU70932, KY270561-KY270659).

NPVL participants were enrolled at the National Neurological AIDS Bank (NNAB), a member of the National NeuroAIDS Tissue Consortium (NNTC)[6] and were on cART until death[7]; it is difficult to obtain such tissues, as the majority of subjects with terminal disease stop taking cART and thus demonstrate a degree of viral rebound. Interestingly, out of the 120 diverse tissues examined in this NPVL cohort, only 24 tissues contained amplifiable HIV *env-nef*, demonstrating the nonubiquitous nature of HIV in tissues during cART, as well as the ability of cART to reduce the anatomical reservoir, but not eliminate the virus completely.[8] By contrast, all 76 tissues from the PVL group contained readily amplifiable HIV. The PVL study group comprised autopsy tissue from patients who died with detectable plasma viral load and AIDS-associated disease pathologies.[9–12]

All DNA sequences in the study were translated into their coded amino acids; and only those with open reading frames were retained for further analysis. For all sequences, variable domain characteristics (length, number of N-sites, and charge) were calculated using the ''Variable Domain Characteristics'' tool at the HIV Sequence Database at Los Alamos with additional calculations performed in Excel (ver. 15.20).

Box and whisker plots appeared to show a trend in different length distributions between PVL and NPVL categories (Fig. 2). Therefore, we used a two-tailed *t*-test (MacWizard ver 1.9.18) to determine whether characteristics of the V1–V5 regions were statistically significantly different between categories at a 95% confidence level. As the number of sequences generated for each subject were uneven, and to avoid biasing the results, in each analysis, the data sets were reduced so that only one sequence of each representative length for each subject was retained for statistical analysis.

There was a significant difference ($p < .01$) in the distributions of length in V1 between NPVL (mean = 30.93 aa's) and PVL (mean = 28.49 aa's) sequences. Furthermore, the range and distribution of amino acid lengths appeared relatively similar among most of the 24 NPVL tissues, with 43% of tissue's sequence populations containing lengths above and below the min/max quartiles of those in the Los Alamos database shown in Figure 1. With exceptions (notably in subjects HCK1 and HCK2), PVL tissues from specific subjects tended to demonstrate less diversity in V1 length, with the V1 sequences from subject's different tissues usually representing either longer or shorter V1s than those in Los Alamos, and only 8% of the PVL tissue's sequence populations spanned both the min/max quartiles observed in the NPVL subject's V1 sequence populations.

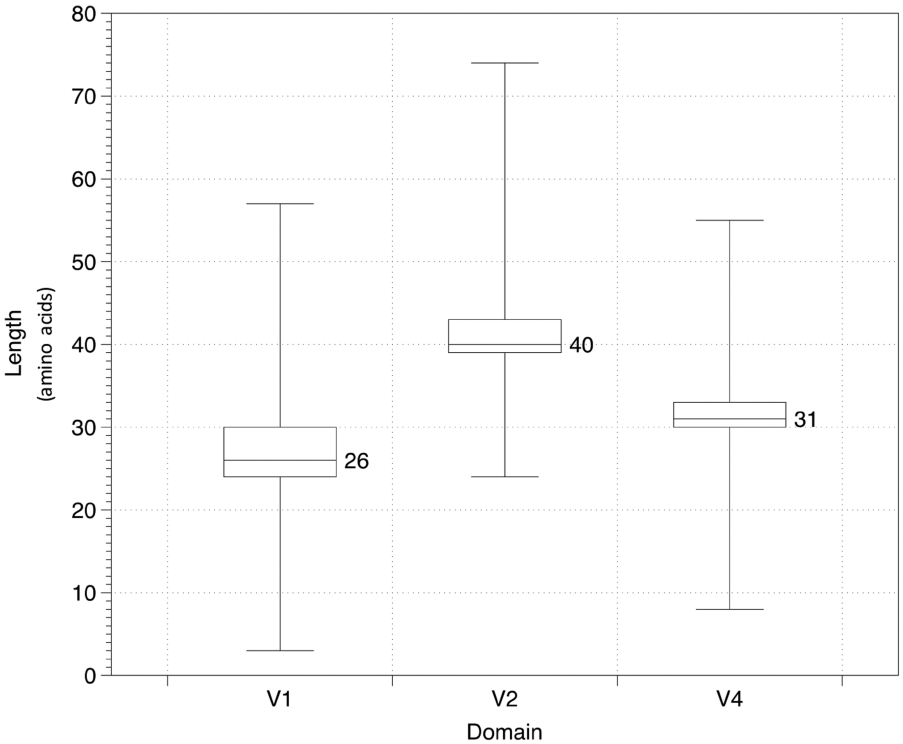

FIG. 1. Length distribution from publicly held V1, V2, and V4 domains at the Los Alamos HIV database. The box spans the interquartile range, with the *top* and *bottom* representing the upper and lower quartiles of the data. The median value is shown and marked by a *vertical line* inside the *box*. The *whiskers* represent the highest and lowest observations.

Table 1. Study Participants

Group	*Sequence approach*	*Subject*	*# Tissues*	*# Sequences*	*VL*	*CD4*	*Therapy history*	*Disease status*
NPVL	SGS	HC02/5095	8	143	<40	58	KTA, TRU	HIVE, plasmacytoma
		HC04/4143	2	23	<400	168	ABC, CBV, TRU, KTA, RTV, DAR, RGV, TMC	KS, lymphoma
		HC05/1010	4	89	<40	154	ddc, 3TC,ZDV,IDV,d4T, RTV, SQV, CBV, NPV, NON, TFV, KTA, ATV	Basal Cell Carcinoma
		HC07/6083	1	18	<40	324	EFV, CBV, NVP, ABC, TFV, KTA, TZV, TFV, 3TC, ETC, TRU	Adenocarcinoma
		HC08/5024	4	40	<40	112	TFV, 3TC, EFP, ddi, KTA, TRU, RTV, DAR	Lung Cancer
		HC09/5025	3	79	<40	268	TFV, ddi, 3TC, KTA	Atherosclerosis, Wasting
		HC10/4124	2	11	<40	48	NFV, 3TC, ZDV,EFV, CBV, NON	Atherosclerosis
PVL	bPCR	AM	7	225	>	UK	UK	AIDS, Lymphoma
		AZ	5	87	>	UK	UK	AIDS, Wasting, Atherosclerosis
		BW	5	121	>	UK	UK	AIDS, Lymphoma
		CX	8	209	>	UK	UK	AIDS, HAD
		DY	7	166	>	UK	UK	AIDS, HAD
		GA	5	132	>	UK	UK	AIDS, HAD
		IV	8	181	>	UK	UK	AIDS, Lymphoma
PVL	SGS	HCK1	3	130	>	UK	UK	AIDS, KS,
		HCK2	3	53	>	UK	UK	AIDS, KS,
		HCK3	3	68	>	UK	UK	AIDS, high-grade metastatic KS
		6,568	2	39	>	77	3TC, ABC, D4T, DDC, DDI, EFV, IDV,NFV, NVP, ZDV	AIDS, HAD, HIVE
		7,766	3	40	1,843	43	3TC, ABC, EFV, D4T, DDI,I DV, NFV, ZDV	AIDS, HAD, HIVE
		10,017	2	29	389,120	7	3TC, D4T, SQV,ZDV	AIDS, HIVE, IVDU
		CA110	4	66	198,957	21	No use reported	AIDS, HAD, HIVE
		NA20	11	20	UK	4	No use reported	Hemophiliac, MCI

Two general groups are listed: NPVL and PVL. For the NPVL group, out of 120 tissues assayed for HIV using a single-genome sequencing approach, only 24 tissues generated HIV *env-nef* sequence data. A " > " character under viral load indicates that we assumed a high viral load based on autopsy reports and the ability to readily amplify HIV from all tissues studied.

UK, unknown; KS, Kaposi's sarcoma; HAD, HIV-associated dementia; HIVE, HIV encephalitis; MCI, mild cognitive disorder; NPVL, no measurable plasma viral load; PVL, plasma viral load. Therapy abbreviations are defined at www.who.int/hiv/pub/guidelines/arv2013/intro/acronyms/en/

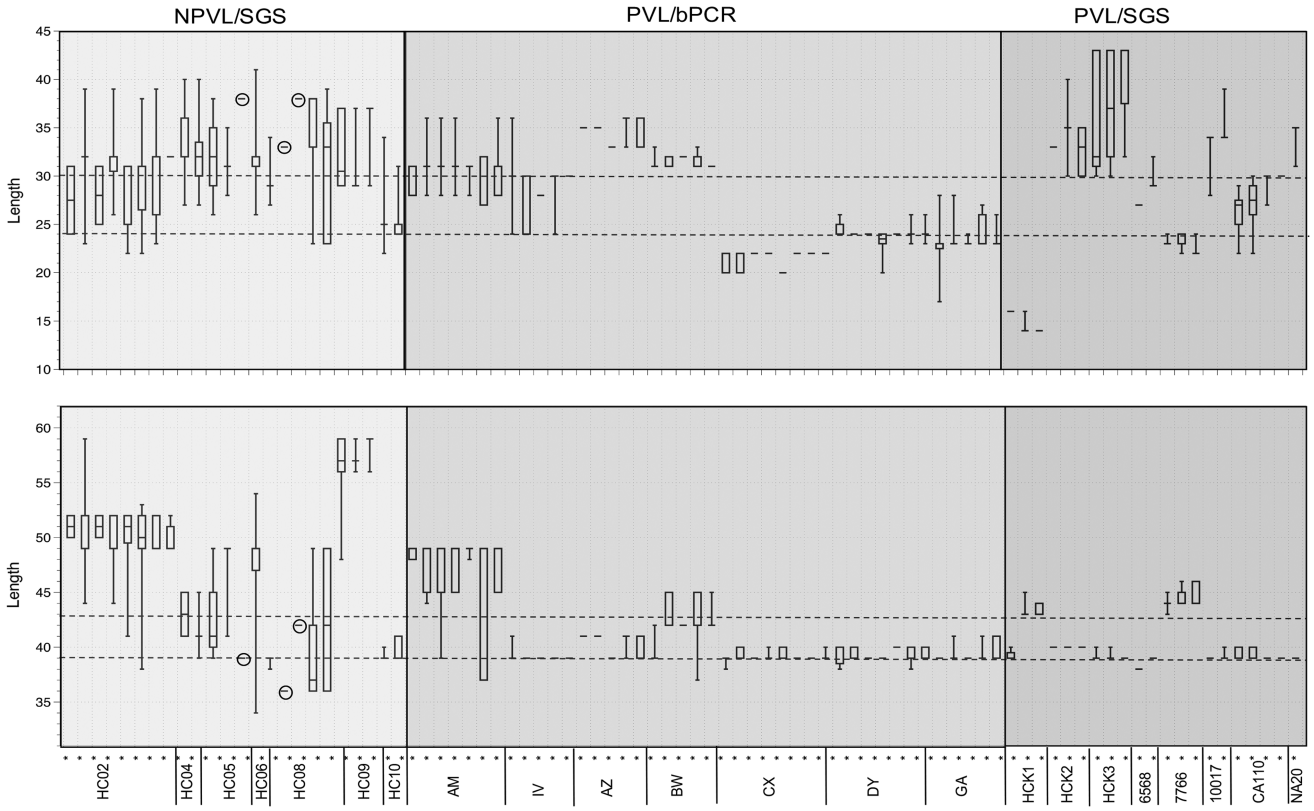

FIG. 2. Env V1 (*top*) and V2 (*bottom*) length variation in individual anatomical tissues. Domain length is on *y*-axis, tissues sampled are on *x*-axis. NPVL = subjects who died with no detectable viral load. PVL = subjects who died with AIDS or a detectable viral load. *Open circles* in NPVL plots represent tissues where only one sequence was obtained. *Dashed horizontal lines* represent the *upper* and *lower* quartiles of the corresponding Los Alamos variable domains. SGS, single-genome sequencing; bPCR, bulk polymerase chain reaction.

In V2, sequences from NPVL subjects tended to be longer in comparison to PVL subjects. Seventy eight percent of the NPVL tissues from 6/7 subjects harbored sequences above the upper quartile identified in the Los Alamos V2 data, whereas only 23% of tissue sequence populations, representing 4/15 PVL subjects, were above the upper quartile in the Los Alamos V2 data. There was no statistical significance to the patterns observed in V3, V4, or V5 domain sequences ($p > .01$)(data not shown).

The length of Env variable domains is known to be highly correlated with an increase in N-sites. In the current study, no statistical difference was observed when comparing the mean number of N-sites between NPVL and PVL datasets in the V1 region ($p > .01$; 2.492 aa's and 2.443 aa's, respectively). In contrast, the mean number of N-sites for the NPVL V2 region was significantly greater than the PVL V2 region ($p < .01$; 3.676 aa's and 2.027 aa's, respectively).

We then assessed variable domains for charge differences by comparing the overall charge of each variable domain and the frequency of positively charged amino acid residues (K and R) and negatively charged amino acid residues (D and E). In V1, NPVL sequences showed a significant increase in the number of charged amino acid residues in comparison to PVL sequences ($p < .01$), whereas no significant difference in the number of charge residues was observed between the groups in the V2 region.

In summary, while previous sequence analysis studies of combined V1–V2 loops noted that longer sequences were associated with an increase in N-sites, in the current analysis, we uncoupled these domains and determined that longer V1 sequences were correlated to an increase in the number of charged residues (both positively and negatively charged), whereas increased glycosylation was associated with longer V2 domains.

We were concerned that the sequencing approach could have impacted our findings. Early studies of HIV primarily used a bulk polymerase chain reaction (bPCR) amplification, wherein many targets were amplified together in the same tube before sequencing. While this method could result in resampling and distorted proportions of viral variants, most studies using bPCR reported highly variable sequence populations. Later, especially when HIV researchers started studying subjects with a very low or undetectable viral load, single-genome sequencing (SGS), which uses endpoint dilution to amplify from only one template in a well, became a more common approach to sequence HIV.

To test to see if these two approaches resulted in statistically different length populations in any variable domains studied, we pooled all bPCR-produced data and all SGS-produced data from the PVL sequence population and performed a two-tailed *t-test*. We found no statistical differences between bPCR and SGS lengths in any variable domain sequence population ($p > .01$).

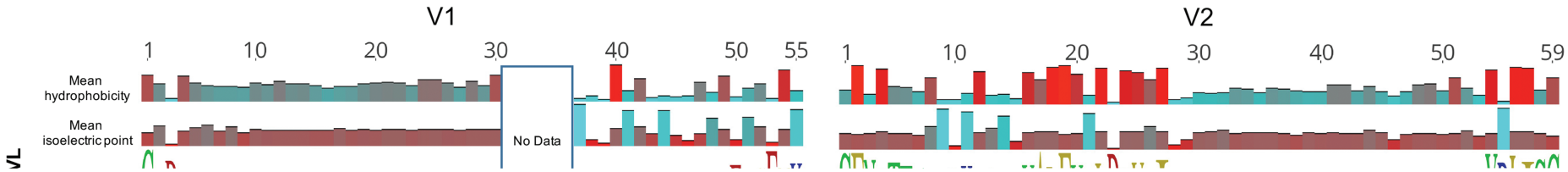

FIG. 3. V1 and V2 amino acid variation in NPVL and PVL sequence alignments. Only nonidentical sequences were used in the alignment (NPVL = 127, PVL = 347 for V1; NPVL = 146, PVL = 399 for V2). Four panels are shown for each alignment: mean hydrophobicity is the hydrophobicity at each residue averaged over all sequences. Mean isoelectric point refers to the pH, at which a molecule carries no net electrical charge. ''SeqLogo'' displays amino acid, where the height of the amino acid is proportional to its frequency at each position in the alignment. ''Identity'' graphically displays the conservation of amino acids across all sequences for every position, where the height of the bar indicates the number of identical amino acids at each position; a *green bar* indicates that the residue at the position is the same across all sequences; *yellow* indicates less than complete identity; *red* indicates very low identity for a given position. Length flexible regions and charged regions are indicated. Color images are available online.

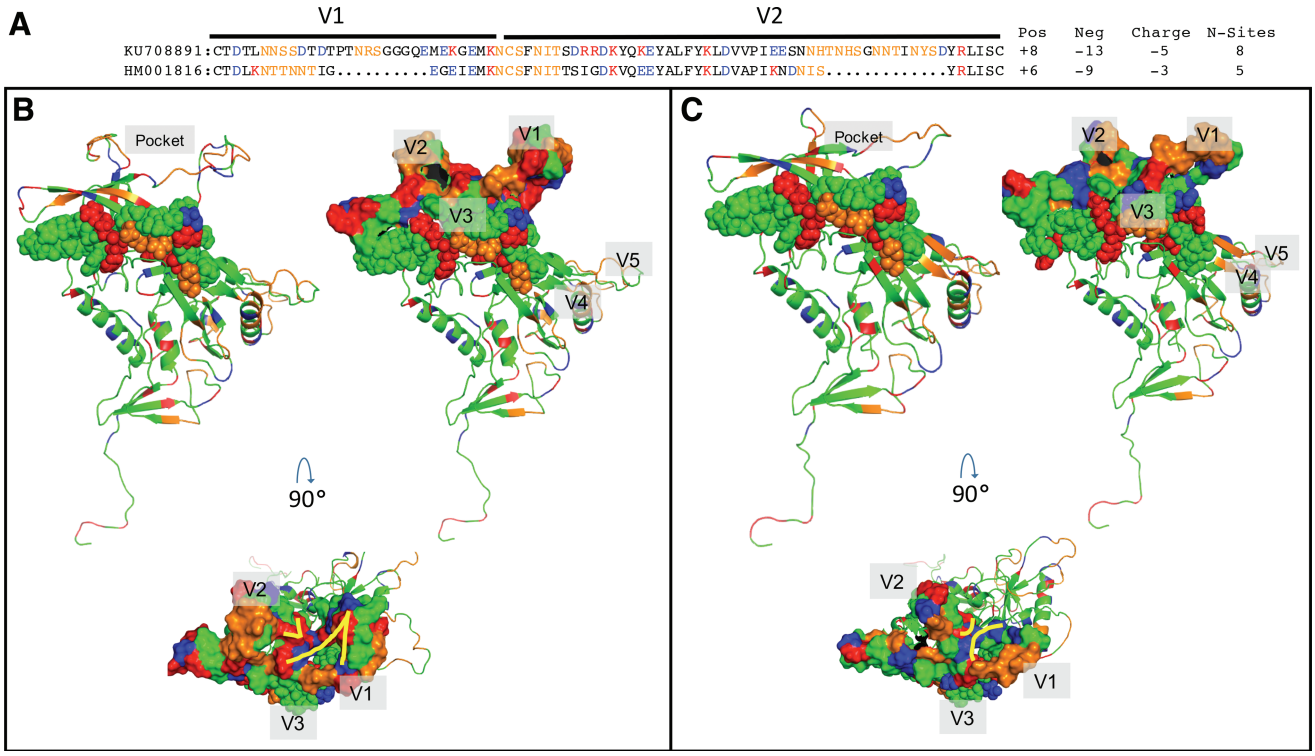

FIG. 5. Structural variation in select long and short V1–V5 sequences. Note that for contrast, we used different sequences than used in Figure 4. In **(A)**, we show the aligned sequences used to develop structures in **(B, C)**, which are colored similarly to Figure 4. In **(B)**, on the *top left*, we show a side view of the long structure. V1–V2 and V4–V5 are shown in cartoon format, and we highlight the V3 domain, which rests below V1/V2 in ''sphere'' format. On the *right*, V1–V2 is shown in a ''surface'' format to demonstrate the complete complex that CD4 must traverse for binding to occur. Below, we have rotated the model 90 degrees to show the top view of the binding pocket, with *yellow lines* highlighting the charged resides in V1 and V2. **(C)** Shows the short structure and the same views as in **(B)**. Note the accumulation of charges (*blue* and *red*) in the pocket of the long V1 structure, compared with the short structure; also note the accumulation of glycosylation sites (*orange*) in the long V2 structure, similar to those observed in Figure 4. The functional progression of CD4 as well as coreceptor binding requires a movement of V1V2 to an ''open'' state allowing for exposure of V3; an extensive description of this process was recently published.[13] Color images are available online.

variable region is similar to V1 in that it comprises many small polar residues that generate putative N-sites. Therefore, for V2, a gap was inserted before the YRLISC motif, and all residues for each sequence were shifted left. Spaces were inserted in the V1 and V2 PVL and NPVL data alignments so that the two alignments (PVL and NPVL) correspond to each other. Alignments and graphics shown in Figure 3 were performed using Geneious software (version 9.1.3).

The SeqLogo consensus graphs (Fig. 3), where the height of the amino acid at each site is proportional to the total numbers of identified amino acids in that column, demonstrate the variability observed within each sequence population. The V1 alignment mirrors results observed in the glycosylation and charge analysis: while both groups have N-sites occurring primarily in the first 20 amino acid residues in the flexible region of glycosylation, the trend of longer sequences in the NPVL group is often due to an abundance of charged positions at the 3′ end of the V1 domain. Also, notable is that the amino acid substitutions in the 3′ charged region can result in considerable differences in hydrophobicity and isoelectric point in the V1 loop.

The V2 SeqLogo graph also clearly demonstrates that the primary difference between NPVL and PVL sequence populations lies in length and in residues that contribute to N-sites (N, T, and S). Despite considerable amino acid variation in both V2 sequence populations, no obvious amino acid substitutions occurred that impacted either the hydrophobicity or isoelectric point along the V2 loop.

Functionally, V3 remains sequestered beneath V1–V2 until a target cell is reached, triggering a conformational rearrangement of V1–V2 to an open state, allowing for exposure of V3.[13] A deeper, highly charged, and glycosylated V3 pocket may mask epitopes recognized by neutralizing antibodies; however, these same variable loop insertions may hinder the conformational process required for CD4 binding.[14] Thus, the modifications observed between long and short V1–V2 domains may have a considerable impact on the function of the envelope.

To observe how length variation in V1 and V2 impacts the structure of the gp120 binding pocket, two sequences representing long V1–V2 domains (GenBank accession numbers KU708975, KU708891) and two sequences representing short V1–V2 sequences (GenBank accession numbers HM002301, HM001816) were chosen for representation as 3D protein structures (Figs. 4 and 5). Structures were modeled using the I-TASSER web server[15,16] and aligned and viewed in MacPyMOL (v1.7.2.1). The top protein threading models were structure 3j70D or 5FYjG from the Protein Data Base (www.rcsb.org/pdb/).

Each pose had a C-score and TM-score within the range of acceptable topologies. In Figure 4, we focus only on the structure of V1–V2. Note that while the overall charge difference between the two sequences is small (−3), the long sequence has twice as many charged residues as the shorter sequence. The structural comparisons clearly demonstrate two other dramatic differences between long and short structures: (1) long sequences have a much deeper binding pocket than short sequences (Fig. 4E), and (2) long sequences have an abundance of charges (contributed by V1) and N-sites (contributed by V2) within the binding pocket (Fig. 4F).

In Figure 5, we characterize two other examples of long and short sequences, this time using the entire V1–V5 structure. Here, the length difference is less than in the structures shown in Figure 4; however, there is still a dramatic contrast between long and short sequences in the depth of the V1 and V2 pocket; note the abundance of charged residues and N-sites residing in the binding pocket (Fig. 5, Top View). In both structures, the V3 loop, which is imperative for CD4 and coreceptor binding, is buried below the V1–V2 complex (Fig. 5A, B).

As a final point, the greater length variation found in HIV Env sequences from tissues of NPVL subjects may have been present but not sampled in the PVL tissues simply due to the abundance of blood-derived virus. As the NPVL subjects were suppressed until death, it is less likely that these tissues were inundated with HIV contained in recently infected circulating T cells. Furthermore, the SGS approach applied to the NPVL tissues would have likely generated data from all viruses present in the tissues, including archived viruses. We hypothesize that for PVL subject's tissues, whether using either SGS or bPCR, the major blood-associated variants may have overshadowed minor and more diversified variants that were observed in the NPVL subject's tissues.

The increased repertoire of Env structures sampled in NPVL tissues, including both long, short, diversely charged, and glycosylated, fosters intriguing hypotheses concerning persistent tissue-based HIV during cART. Any hypothesis must take into account that tissue-based HIV during cART likely contains archival virus integrated in immune cells, which might be hypermutated or nonfunctional.

One hypothesis suggests that as Env lengths fluctuate over time in blood, the variants repeatedly infect tissue-based immune cells and remain over an extensive period of time with limited evolution, thus the diversity observed in tissue-based HIV is primarily due to its historical nature.

Another hypothesis suggests that tissue-based HIV represents subpopulations of viable and compartmentalized virus that continues to replicate in response to a less understood and complex tissue-based microenvironment. Furthermore, these subpopulations may be masked in subjects who are not on cART due to a massive number of recently replicated viruses.

A third hypothesis combines both ideas, suggesting that the findings presented here are consistent with tissue-based virus populations with both archival and replicating variants. Importantly, functional and longer tissue-based forms of gp120 may be more effective in evading the immune system, as these are tightly linked to immune viral escape and chronic infection[3,14] and thus require attention for consideration as vaccine targets. At the same time, HIV containing functional and shorter Envs in tissues during cART may represent a population of viruses that are more easily transmitted[3,14] or could more easily infect circulating T cells in the absence of cART and thus may play a role in viral rebound post-cART. Thus, NPVL tissue-derived Env sequences, while smaller in numbers than those in PVL tissues, may provide essential mechanistic insights into viral persistence and immune escape that deserves more attention.

As cART has been unable to cure HIV infection, the study of HIV persistence and evolution in tissues is of increased interest. Most studies concerning HIV diversity under cART primarily consider the genes that cART directly targets (e.g., pol, integrase). Remarkably, this study suggests that studies based on Env data derived from tissues of PVL patients may not observe genetic variation representative of an ongoing and dynamic interplay between resident immune cells and tissue-based HIV.

This study also highlights the importance in tissue selection and patient cART history to fully understand tissue-based HIV populations. Using a combination of SGS and deep sequencing approaches to amplify HIV in tissues, such as ultradeep single-molecule real-time sequencing which generates thousands of full-length HIV gene sequences,[17] is extremely useful as the approach can generate long gp120 sequences that reveal the complete sequence Env population in a tissue, including minor variants.

Acknowledgments

We acknowledge the following tissue banks and especially their participants and donors: the AIDs and Cancer Specimen Resource, the National Neurological Database, the National NeuroAIDS Tissue Consortium, and the Medical Research Council Brain and Tissue Bank (Edinburgh, UK). This study was funded by the following NIH mechanisms: NIMH R01-MH100984, NINDS R01-NS095749, NINDS R01-NS107022, and NIMH/NINDS U24MH100929

Author Disclosure Statement

No competing financial interests exist.

References

1. McLellan JS, Pancera M, Carrico C, *et al.*: Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 2011;480:336–343.
2. Moore PL, Gray ES, Wibmer CK, *et al.*: Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. Nat Med 2012;18:1688–1692.
3. 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;79:6528–6531.
4. Frost SD, Liu Y, Pond SL, *et al.*: Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. J Virol 2005;79:6523–6527.
5. Salazar-Gonzalez JF, Salazar MG, Keele BF, *et al.*: Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med 2009;206:1273–1289.
6. Morgello S, Gelman BB, Kozlowski PB, *et al.*: The National NeuroAIDS Tissue Consortium: A new paradigm in brain banking with an emphasis on infectious disease. Neuropathol Appl Neurobiol 2001;27:326–335.

7. Lamers SL, Rose R, Maidji E, *et al.*: HIV DNA is frequently present within pathologic tissues evaluated at autopsy from cART-treated patients with undetectable viral load. J Virol 2016;90:8968–8983.
8. Nolan DJ, Rose R, Rodriguez PH, *et al.*: The Spleen Is an HIV-1 Sanctuary During Combined Antiretroviral Therapy. AIDS Res Hum Retroviruses 2018;34:123–125.
9. Ayers LW, Silver S, Orenstein JM, McGrath MS, Garcia DL: The AIDS and Cancer Specimen Resource. Methods Mol Biol 2011;675:193–203.
10. Lamers SL, Salemi M, Galligan DC, *et al.*: Extensive HIV-1 intra-host recombination is common in tissues with abnormal histopathology. PLoS One 2009;4:e5065.
11. Lamers SL, Rose R, Nolan DJ, *et al.*: HIV-1 Evolutionary Patterns Associated with Metastatic Kaposi's Sarcoma during AIDS. Sarcoma 2016;2016:4510483.
12. Gonzalez-Perez MP, Peters PJ, O'Connell O, *et al.*: Identification of Emerging Macrophage-Tropic HIV-1 R5 Variants in Brain Tissue of AIDS patients without severe neurological complications. J Virol 2017;91:e00755–17.
13. Ozorowski G, Pallesen J, de Val N, *et al.*: Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. Nature 2017;547:360–363.
14. Gnanakaran S, Bhattacharya T, Daniels M, *et al.*: Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. PLoS Pathog 2011;7:e1002209.
15. Roy A, Kucukural A, Zhang Y: I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–738.
16. Zhang Y: I-TASSER server for protein 3D structure prediction. BMC Bioinform 2008;9:40.
17. Brese RL, Gonzalez-Perez MP, Koch M, *et al.*: Ultradeep single-molecule real-time sequencing of HIV envelope reveals complete compartmentalization of highly macrophage-tropic R5 proviral variants in brain and CXCR4-using variants in immune and peripheral tissues. J Neurovirol 2018; 24:439–453.

Address correspondence to:
Susanna L. Lamers
BioInfoExperts LLC
718 Bayou Lane
Thibodaux, LA 70301

E-mail: susanna@bioinfox.com