

Research Article

Evidence of Divergent Amino Acid Usage in Comparative Analyses of R5- and X4-Associated HIV-1 Vpr Sequences

Gregory C. Antell,^{1,2,3} Will Dampier,^{1,2,3} Benjamas Aiamkitsumrit,^{1,2} Michael R. Nonnemacher,^{1,2} Vanessa Pirrone,^{1,2,4} Wen Zhong,^{1,2} Katherine Kercher,^{1,2} Shendra Passic,^{1,2} Jean Williams,^{1,2} Yucheng Liu,^{1,2} Tony James,^{1,2} Jeffrey M. Jacobson,^{4,5} Zsofia Szep,^{4,5} Brian Wigdahl,^{1,2,4,6} and Fred C. Krebs^{1,2}

¹Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, USA

²Center for Molecular Virology and Translational Neuroscience, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, PA, USA

³School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA, USA

⁴Center for Clinical and Translational Medicine, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, PA, USA

⁵Division of Infectious Diseases and HIV Medicine, Department of Medicine, Drexel University College of Medicine, Philadelphia, PA, USA

⁶Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA

Correspondence should be addressed to Fred C. Krebs; fkrebs@drexelmed.edu

Received 10 November 2016; Accepted 20 March 2017; Published 17 May 2017

Academic Editor: Margarita Hadzopoulou-Cladaras

Copyright © 2017 Gregory C. Antell et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vpr is an HIV-1 accessory protein that plays numerous roles during viral replication, and some of which are cell type dependent. To test the hypothesis that HIV-1 tropism extends beyond the envelope into the *vpr* gene, studies were performed to identify the associations between coreceptor usage and Vpr variation in HIV-1-infected patients. Colinear HIV-1 Env-V3 and Vpr amino acid sequences were obtained from the LANL HIV-1 sequence database and from well-suppressed patients in the Drexel/ Temple Medicine CNS AIDS Research and Eradication Study (CARES) Cohort. Genotypic classification of Env-V3 sequences as X4 (CXCR4-utilizing) or R5 (CCR5-utilizing) was used to group colinear Vpr sequences. To reveal the sequences associated with a specific coreceptor usage genotype, Vpr amino acid sequences were assessed for amino acid diversity and Jensen-Shannon divergence between the two groups. Five amino acid alphabets were used to comprehensively examine the impact of amino acid substitutions involving side chains with similar physicochemical properties. Positions 36, 37, 41, 89, and 96 of Vpr were characterized by statistically significant divergence across multiple alphabets when X4 and R5 sequence groups were compared. In addition, consensus amino acid switches were found at positions 37 and 41 in comparisons of the R5 and X4 sequence populations. These results suggest an evolutionary link between Vpr and gp120 in HIV-1-infected patients.

1. Introduction

HIV-1 entry proceeds via direct interaction between the viral envelope glycoprotein (gp) 120 and the host cell CD4 receptor molecule, as well as one of two coreceptor molecules, CCR5 or CXCR4 [1]. The HIV-1 gp120 envelope protein consists of five hypervariable regions (V1–V5) interspersed

among five constant regions (C1–C5), with the third variable loop (V3) region serving as the major determinant in HIV-1 coreceptor phenotype. Specifically, V3 consists of 34–36 amino acid residues (most commonly 35 residues). However, as conformational changes within the V3 loop occur upon the binding of gp120 with CD4, it is possible that the coreceptor usage-associated amino acid residues within other

regions of envelope participate in the structural rearrangement of gp120 [2, 3]. Patients with HIV-1 infections that are not well suppressed by effective antiretroviral therapy will experience some form of coreceptor switch, in which CXCR4-utilizing (X4) viruses gradually emerge from a predominantly CCR5-utilizing (R5) swarm due to the accumulation of amino acid changes within the V3 loop, particularly at amino acid positions 11 and 25. HIV-1 tropism based on coreceptor usage was previously used interchangeably with tropism defined by cellular target. However, recent studies have concluded that it is more precise to discuss viral infection in terms of utilization of a coreceptor alongside the phenotype of the target cell [4, 5]. For genomic applications, coreceptor utilization prediction algorithms can be effectively used to classify Env-V3 sequences computationally, as previously reviewed [6].

Variations in Env-V3 sequences that alter coreceptor phenotype are specific examples of sequence changes that take place throughout the HIV-1 genome during an ongoing HIV-1 replication. The present studies were based on the general hypothesis that changes in Env-V3 that alter the coreceptor phenotype lead to changes elsewhere in the genome. This is likely due to functional changes that are linked to coreceptor usage and, by extension, the identity of the infected cell. The overall goal of this study was to perform comparative analyses of the accessory viral protein R (Vpr) with regard to R5 and X4 HIV-1 quasiespecies. Vpr is a 96 amino acid protein with three characteristic α -helical secondary structures, α -helix 1 (positions 17–33), α -helix 2 (positions 38–50), and α -helix 3 (positions 53–78). The three helices are flanked by two, nonstructural, but functional, N- and C-terminal domains [7–9]. In general, Vpr has been understudied relative to other HIV-1 proteins, despite its multifaceted roles in the replication cycle, activity as an extracellular protein, and contributions to HIV-1-associated neuropathogenesis. Most critically, Vpr is essential for viral replication in macrophages but dispensable in T cells, suggesting cell type-specific functionality and possible coevolution with Env-V3 [10]. Since coreceptor usage is an important (but not absolute) determinant of cellular tropism, we hypothesized that coreceptor-associated genetic signatures would emerge in Vpr due to cell type-specific selective pressures that similarly direct coreceptor usage. Support for this hypothesis was provided by previous reports of specific HIV-1 Tat and LTR sequence signatures associated with coreceptor utilization, suggesting that selective pressures applied within specific cellular targets during the course of HIV-1 infection result in distinct transcriptional regulatory mechanisms [3, 6, 11]. Although coreceptor usage has been clearly linked to HIV-1 transmission as well as HIV-1-associated disease progression [12–15], the degree and specificity of variations in Vpr that might contribute to such clinically significant events remain unexplored.

In this study, we extended a genetic approach previously applied to investigate gp120, Tat, and LTR associations with coreceptor utilization [3] to the accessory protein Vpr. Aligned Vpr sequences were grouped as X4 or R5 according to the predicted coreceptor utilization of the colinear V3 region and evaluated independently at each Vpr amino acid

position with regard to amino acid diversity and Jensen-Shannon divergence between the groups. Combined diversity and divergence metrics permitted the quantitative characterization of position-specific amino acid variation and provided a methodology for comparing sequence information between different populations. Overall, this strategy provided a concise genetic approach that revealed specific amino acid positions and side chain motifs in Vpr that were coselected with CXCR4 or CCR5 coreceptor utilization during replication in HIV-1-infected patients.

2. Materials and Methods

2.1. Collection and Sequencing of DNA from HIV-1-Infected Patients. Patients enrolled in the Drexel Medicine CNS AIDS Research and Eradication Study (CARES) Cohort were recruited from the Partnership Comprehensive Care Practice of the Division of Infectious Diseases and HIV Medicine in the Department of Medicine at Drexel University College of Medicine (Philadelphia, Pennsylvania, USA) and the Center for Clinical and Translational Medicine in the Institute for Molecular Medicine and Infectious Disease. Patients in the Drexel Medicine CARES Cohort were recruited under protocol 1201000748 (Brian Wigdahl, PI), which was approved by the Drexel University College of Medicine Institutional Review Board (IRB) and adheres to the ethical standards of the Helsinki Declaration (1964, amended most recently in 2008) developed by the World Medical Association. All patient samples were collected under the auspices of this protocol through written consent obtained upon patient enrollment. Whole blood collected from each patient was processed and assessed as previously described [16]. Peripheral blood samples were used for drug screening, plasma analysis, and PBMC isolation as previously described. Env-V3 and Vpr regions were sequenced from 4.4 kb fragments [11, 16].

2.2. Compiling Colinear HIV-1 Vpr-Env Sequences. HIV-1 sequences from the Los Alamos National Laboratory (LANL) HIV Sequence Database that included colinear *env*-V3- and *vpr*-coding regions were collected and annotated in October 2014, while additional HIV-1 sequences from patient samples collected from the Drexel Medicine CARES Cohort were included to supplement the publicly available dataset. The Drexel Medicine CARES Cohort is a predominantly subtype B patient population from Philadelphia, Pennsylvania [11, 16–18]. To reduce the impact of sequence variation due to regional differences, the query of the LANL sequence database was limited to include only North American sequences. Furthermore, the search query was restricted to one sequence per patient and excluded laboratory strain sequences or those used for previous functional studies. Vpr sequences added from the Drexel Medicine CARES Cohort have been submitted to GenBank under BioProject sub1322660.34.

2.3. Coreceptor Usage Classification. In silico, coreceptor usage predictions were performed using WebPSSM on Env-V3 amino acid sequences colinear with the Vpr sequences.

Sequences were classified as CCR5 or CXCR4 utilizing (R5 or X4) [19]. This algorithm classifies CXCR4 coreceptor utilization at both high sensitivity (84%) and specificity (96%) [19]. Exclusion methods were applied to reduce noise introduced by WebPSSM predictions as previously discussed [20]. Specifically, V3 sequences were required to be exactly 35 amino acids in length, passed the standard quality control metrics implemented by WebPSSM, and have a WebPSSM score outside the “indeterminate range” (defined as scores between -2.88 and -6.96). With the application of this filtering method, the genetic analyses were focused on sequences with the greatest confidence classification in the PSSM-derived distribution, resulting in the definitive characterization of Env-V3 sequences as CCR5- or CXCR4-utilizing genotypes [11].

2.4. Reduced Amino Acid Alphabets. Reduced amino acid alphabets were utilized in order to gain deeper insight into the structural and functional consequences of amino acid substitutions. The 20-letter amino acid alphabet was reduced to smaller alphabets based on correlations indicated by the BLOSUM50 similarity matrix [21] and implemented in the Biopython version 1.66 [22]. The resulting alphabets were designated as Murphy-15, Murphy-10, Murphy-8, and Murphy-4, referring to the number of letters that make up each alphabet. Accordingly, all diversity and divergence calculations were performed on reduced amino acid alphabets in addition to the full alphabet containing 20 distinct amino acids. Whereas diversity and divergence metrics may be over-sensitive to amino acid variation between amino acids with similar physiochemical properties (K versus R for example), reduced alphabets may overgeneralize and exclude important sequence information. Overall, this methodology aims to better focus on the truly important regions of a protein by observing trends in diversity and divergence across multiple alphabets.

2.5. Positional Diversity Analysis. Vpr sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE), (version 5.05) [23] using default parameters, with insertions and gaps removed to preserve reference numbering relative to the HXB2 reference sequence (K03455). The amino acid diversity of each position was calculated using a window length, w , of 1 and an order of 1 [equivalent to exp (Shannon entropy with base e)] according to the following equation [24]:

$$D_{w,p} = \exp \left(- \sum_{i=1}^{R_{w,p}} P_{i,w,p} \ln [P_{i,w,p}] \right). \quad (1)$$

In this definition, genetic diversity, D , is defined as a weighted sum of all variants, p , at a given position, i , in the amino acid sequence. While any window length, w , may be applied, $w=1$ was used to independently assess diversity at each position within a multiple sequence alignment. The amino acid diversity calculates the effective number of amino acids present at each alignment position and gives greater weight to neither rare nor abundant species when the order=1. Generally, low-diversity positions have a higher

likelihood of exhibiting structural and functional importance, while positions more permissive to structural and functional variation display higher diversity at the amino acid level [25].

2.6. Positional Jensen-Shannon Divergence Analysis. Jensen-Shannon divergence, which is based on Kullback-Leibler divergence, measures the similarity between two probability distributions. This analysis has been successfully applied to previous profile-to-profile multiple sequence alignment comparisons [26, 27]. Jensen-Shannon divergence scores are bound in the range of 0 (completely similar) and 1 (completely dissimilar), with dissimilarity identifying positions of interest. Position frequency matrices (PFMs) were generated from multiple sequence alignments (MSA) of R5- and X4-classified sequence populations, with each PFM containing the relative abundance for each potential residue (or residue group in the case of reduced alphabets) across all positions. PFMs derived from R5- and X4-classified Vpr sequences were used to compute the Jensen-Shannon divergence between populations using the following equation [3]:

$$D_{js} = \frac{1}{2} \left[\sum_{a=1}^{20} Q_a^1 \log_2 \frac{Q_a^1}{Q_a^0} + \sum_{a=1}^{20} Q_a^2 \log_2 \frac{Q_a^2}{Q_a^0} \right], \quad (2)$$

where $Q_a^0 = 1/2(Q_a^1 + Q_a^2)$.

2.7. Statistical Analysis. Statistical analyses were completed in custom IPython Notebooks using the SciPy Python library (version 0.14.0). To identify statistically significant positions, a Monte Carlo permutation test ($n = 1000$) was applied for groups of size M and N , where M and N represent the number of sequences in the X4 and R5 Vpr groups, respectively. Numerical integration of the probability density function (PDF) of the randomly generated Jensen-Shannon divergence distribution was implemented using a Gaussian kernel density estimator in SciPy. P values were determined according to the probability of arriving at a random value equal to or greater than the true Jensen-Shannon divergence.

3. Results and Discussion

Studies of coordinated amino acid variation between Vpr and gp120 were initiated using HIV-1 DNA sequences containing both *vpr* and *env*. Sequences of clinical origin were derived from blood samples collected from HIV-1-infected patients in the Drexel Medicine CNS AIDS Research and Eradication Study (CARES) Cohort [11, 16–18, 28, 29]. Colinear DNA was isolated from patient PBMCs and amplified using primers specifically designed to encompass the Vpr- and gp120-coding sequences. Similarly, DNA sequences from the Los Alamos National Laboratory (LANL) HIV-1 Databases (www.hiv.lanl.gov) were selected based on the availability of both reading frames in a single, colinear sequence. Sequences derived from patient DNA and acquired from the LANL Databases were first analyzed to determine the putative coreceptor phenotype of the encoded gp120 glycoprotein. Sequences were analyzed using WebPSSM, an established and recognized online

TABLE 1: Identification of HIV-1 Vpr sequences colinear to Env-V3 sequences defined as R5 or X4. A total of 1010 Vpr sequences were classified according to coreceptor utilization using WebPSSM [19].

Source	CCR5	CXCR4
LANL	949	32
CARES	20	9
Total	969	41

tool for predicting coreceptor usage from amino acids encoded by HIV-1 gp120 V3 sequences [19]. Of the 1010 colinear sequences included in the final analysis, 969 were found to contain V3 sequences that scored as R5 genotypes. In contrast, only 41 sequences were scored as X4 genotypes (Table 1). The total set of sequences analyzed included 29 samples from the Drexel CARES Cohort, 20 of which were classified as R5.

3.1. HIV-1 Vpr Amino Acid Diversity Is Concentrated in Two Major Regions of Vpr. In order to quantitatively assess the meaningful amount of variation occurring within R5- and X4-associated Vpr sequences at the amino acid level, the amino acid diversity for each Vpr position was independently calculated. Furthermore, a series of reduced amino acid alphabets (Murphy-15, Murphy-10, Murphy-8, and Murphy-4) were utilized to assess amino acid diversity [21]. These alphabets hierarchically group amino acids with similar physiochemical properties (Figure 1). As a general trend, the magnitude of the greatest diversity values decreased in concert with reduction of the amino acid alphabet size in both the R5 and X4 Vpr populations. However, the most diverse positions tended to be robustly persistent against the reduction in alphabet size for all alphabets except the highly generalized Murphy-4 alphabet (Figure 2).

A diversity value of 2 in either the R5 or X4 group, using the standard 20 amino acid alphabet, was used as a threshold to define highly diverse positions. The identified positions were located at residues 28, 37, 41, 45, 55, 77, 84, 85, 89, and 93 (Table 2), spanning all three major helical structures of Vpr. Overall, there was no clear trend for either the X4 or R5 populations to consistently display greater diversity across all positions, and most positions displayed approximately equal diversity in the respective populations. In almost all cases, Vpr amino acid diversity was reduced to less than 2 when using the Murphy-4 alphabet, indicating the strong propensity for amino acid substitutions to be contained within the same physiochemical groups once the group sizes become sufficiently large.

There were, however, noteworthy diversity trends at four positions (28, 37, 55, and 89) that persisted across multiple alphabets. Position 28 displayed considerably greater diversity in X4 Vpr when using the full and Murphy-15 alphabets (4.45 versus 3.37 in both cases), but the difference narrowed considerably using the Murphy-10 and Murphy-8 alphabets (3.18 versus 2.95 in both cases). This difference was primarily due to the R5 population being dominated by three variants (R28, N28,

and S28) while the X4 population displayed a more even distribution. Positions 55 and 89 both displayed similar trends, in which the R5 diversity trended higher (and over the diversity level threshold of 2) across all alphabets but Murphy-4. This was largely due to the presence of an E55 variant in the R5 population that was entirely absent among X4 Vpr sequences. Similarly, the differential diversity of position 89 was attributed to variant A89 being more highly conserved in the X4 population, with an A89T substitution having no influence on diversity in the Murphy-4 alphabet. Finally, position 37, which was extremely polymorphic, was unique in that the more diverse group switched with the amino acid alphabet utilized. While the X4 group appeared to be more diverse using the standard 20 amino acid alphabets, R5 diversity was greater for each of the four reduced alphabets. At position 37, the most common X4 variants L37, V37, I37, and M37, which contributed to the diversity at this position, collapsed into one group in the Murphy-15 alphabet, while common R5 variants such as P37, S37, and E37 persisted in distinct amino acid groups as the alphabet was reduced.

3.2. Divergence in Amino Acid Utilization between X4 and R5 Vpr Is Highly Consistent across Reduced Amino Acid Alphabets. Jensen-Shannon divergence was calculated between amino acid distributions at each position in R5- and X4-associated Vpr sequences (Figure 3), using both the standard amino acid alphabet and reduced alphabets (Figure 1). Jensen-Shannon divergence identifies positions with differential amino acid profiles between R5 and X4 Vpr sequences. Statistically divergent positions ($P < 0.01$) were highly consistent across all alphabets, indicating that the amino acid variation driving Jensen-Shannon divergence was maintained across amino acid categories with increasingly generalized physiochemical properties. Vpr positions 36, 37, 41, 89, and 96 were identified in multiple levels of analysis (Table 3). These positions are concentrated near the second alpha helix of Vpr (36, 37, and 41) and the C-terminal domain (89 and 96).

3.3. Amino Acid Substitution Biases in Vpr Are Associated with Coreceptor Genotype. Sequence logos were utilized to visualize the two regions of HIV-1 Vpr where amino acid positions displayed significant Jensen-Shannon divergence between X4 and R5 Vpr sequences (Figure 4). In these analyses, Vpr regions 36–45 and 87–96 were selected to encompass all five positions with statistically significant Jensen-Shannon divergence scores in at least one alphabet (Table 3). Positions 37, 41, and 89 displayed consistently high and statistically significant Jensen-Shannon divergence. In accordance with large Jensen-Shannon divergence, positions 37 and 41 both featured considerable shifts in amino acid frequencies and differences in consensus between the R5 and X4 populations, with R5 featuring I37/S41 and X4 featuring P37/G41. In contrast, position 45 was highly polymorphic (H45 and Y45 and a diversity greater than 2 regardless of alphabet) but was not divergent with respect to predicted coreceptor phenotype. Although significant divergence was observed in the

AA	M15	M10	M8	M4
L	L	L	L	L
V	L	L	L	L
I	L	L	L	L
M	L	L	L	L
C	C	C	L	L
A	A	A	A	A
G	G	G	A	A
S	S	S	S	A
T	T	S	S	A
P	P	P	P	A
F	F	F	F	F
Y	F	F	F	F
W	W	F	F	F
E	E	E	E	E
D	D	E	E	E
N	N	E	E	E
Q	Q	E	E	E
K	K	K	K	E
R	K	K	K	E
H	H	H	H	E

FIGURE 1: Reduced alphabets utilized for analysis of Vpr amino acid variation. The 20 standard amino acids were grouped according to BLOSUM50 correlations in a hierarchical manner as previously reported in [21]. The resulting alphabets were designated as Murphy-15, Murphy-10, Murphy-8, and Murphy-4. Colors represent amino acids grouped in each alphabet according to physiochemical properties (M15, M10, M8, and M4, resp.).

Murphy-4 alphabet at position 36 (with the introduction of W36 in the X4 sequences), the consensus at this position (R36) was the same in X4 and R5 sequences. The significant coreceptor phenotype-dependent divergence and consensus changes at positions 37 and 41 provide some evidence for the coordinated or parallel evolution of coding sequences in *vpr* and the V3 portion of *env*.

Similar changes, however, were not observed in conjunction with significant divergence at positions in the C-terminal region of Vpr. At position 89, significant Jensen-Shannon divergence was not accompanied by a consensus shift, although a trend toward a shift from T89 in the X4 sequences to A89 in the R5 sequences was evident. At position 96, the consensus remained S96 even with significant divergence in four of the five alphabets (Table 3). In another example of the independence between divergence and consensus, position 77 displayed a consensus difference between the X4 (Q77) and R5 (R77) populations despite the absence of Jensen-Shannon divergence at this position (data not shown). These results may reflect the lesser importance of specific residues in the C-terminal portion of Vpr with respect to coreceptor usage.

4. Conclusions

The role that genetic variation beyond the HIV-1 envelope plays in viral tropism and replication remains a largely unexplored domain. We hypothesized that Vpr would be subject to differential sequence variation as a consequence of cell type-specific roles played by Vpr during replication in a target cell determined by viral coreceptor usage. HIV-1 Vpr has been shown to exist in cell-associated, virion-associated, and virus-free extracellular forms, implicating the protein in multiple aspects of the HIV-1 life cycle [30]. In particular, the virion-associated form of Vpr points to a role for the protein in the early stages of HIV-1 entry and infection, perhaps in the later stages of the viral life cycle as progeny virions are assembled, and with respect to the structural integrity of the final infectious virion, at least with virus produced within cells of the monocyte-macrophage lineage [31]. Vpr associated with virions enables the translocation of the preintegration complex (PIC) into the nucleus of cells such as macrophages [32]. As such, there is also a strong requirement for Vpr during HIV-1 replication in monocytes-

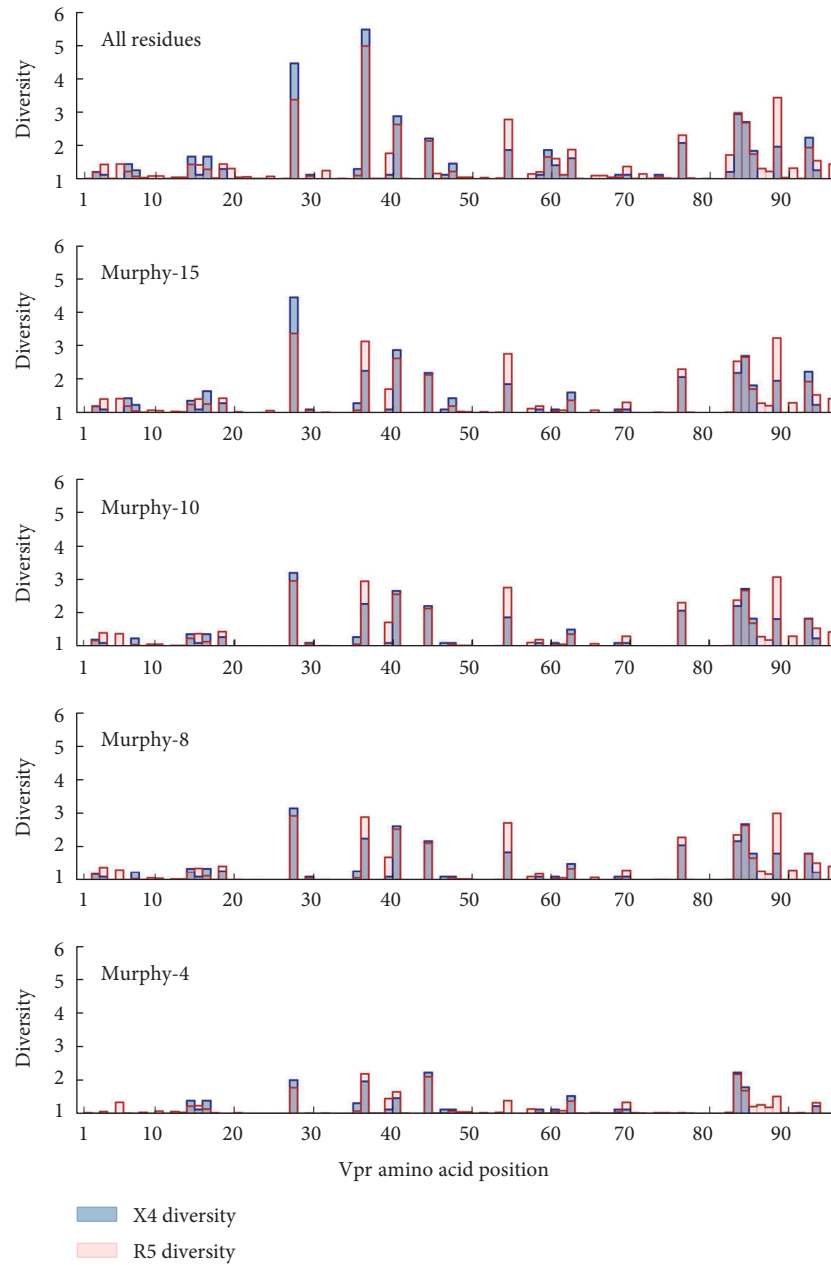


FIGURE 2: Position-specific Vpr amino acid diversity is generally consistent across multiple amino acid alphabets but different between R5- and X4-associated sequences. The amino acid diversity was calculated across all positions of Vpr for both the R5 (red) and X4 (blue) sequence populations. While the overall levels of variation tended to be consistent between the two populations, there were notable differences in amino acid diversity between the R5 and X4 populations that persisted across multiple alphabets at positions 28, 37, 55, and 89. Diversity values for amino acid positions at which at least one diversity value was greater than 2 in the full amino acid alphabet are listed in Table 2.

macrophages, but expendability in T lymphocytes [10, 33]. In these varied roles, Vpr might be subjected to cell-dependent pressures that would impact the emergence of variants dependent on coreceptor usage and, by extension, host cell type.

The combined use of diversity and divergence analysis allowed us not only to survey the total amount of variation in Vpr but also to test the hypothesis that variation is different between coreceptor-defined sequence groups at specific positions. As such, this methodology allows us to independently assess differences at each amino acid

position that may be under selection, a result not afforded by a standard phylogenetic analysis. Due to the presence of high diversity at many amino acid positions throughout Vpr (Figure 2), observation of consensus changes may not always be statistically meaningful with respect to Jensen-Shannon divergence. Alternatively, it is also possible for significantly divergent positions to not demonstrate a consensus change in Vpr amino acid usage between R5 and X4. This is because a change in the relative presence of variants in a polymorphic position, although an indicative of a consensus switch, may fall within the limits of

TABLE 2: Diversity values at amino acid positions in Vpr. Each position in the table had at least one diversity value greater than 2 in the full amino acid alphabet. A diversity value greater than 2 (bold) indicates a diverse amino acid position.

Position	Coreceptor	All residues	Murphy-15	Murphy-10	Murphy-8	Murphy-4
28	X4	4.45	4.45	3.18	3.18	1.99
	R5	3.37	3.37	2.95	2.95	1.76
37	X4	5.46	2.27	2.27	2.27	1.94
	R5	4.97	3.14	2.93	2.92	2.16
41	X4	2.88	2.88	2.65	2.65	1.45
	R5	2.63	2.63	2.55	2.55	1.64
45	X4	2.20	2.20	2.20	2.20	2.20
	R5	2.14	2.14	2.14	2.14	2.08
55	X4	1.87	1.87	1.87	1.87	1.00
	R5	2.78	2.76	2.74	2.74	1.38
77	X4	2.07	2.07	2.07	2.07	1.00
	R5	2.31	2.31	2.31	2.31	1.02
84	X4	2.94	2.20	2.20	2.20	2.20
	R5	2.97	2.55	2.38	2.38	2.15
85	X4	2.71	2.71	2.71	2.71	1.78
	R5	2.68	2.66	2.66	2.66	1.68
89	X4	1.97	1.97	1.83	1.83	1.00
	R5	3.42	3.24	3.06	3.02	1.50
93	X4	2.23	2.23	1.83	1.83	1.00
	R5	1.94	1.94	1.82	1.82	1.00

uncertainty if the magnitude of the change is modest. In order to better focus our understanding of amino acid substitutions on potentially functional changes, we expanded our analysis to include a series of reduced amino acid alphabets. While larger alphabets preserve more information from the original amino acid sequence, they can also overemphasize variation among amino acids with similar physiochemical properties. Alternatively, although reduced alphabets exclude potentially meaningful information contained in the original sequence, they maintain differences between physiochemically disparate residues.

The functional impact of the specific amino acid substitutions identified in this study is difficult to infer precisely in the absence of specific mutagenic studies. Nevertheless, positions with differential amino acid occupancy in X4 and R5 sequences were found in regions of Vpr functionally associated with viral packaging, G₂ arrest, nuclear localization, and alterations in disease severity or progression [34–37]. Of the specific positions observed in this study, positions 36 and 37 are regarded as essential for Vpr oligomerization [38] while position 89 is possibly correlated with clinical manifestations of HIV-1 disease [39]. Structurally, both glycine and proline substitutions may play large roles in altering Vpr secondary structure, especially in alpha helices. Position 37, which is occupied by a proline in consensus R5 Vpr, could potentially lead to the formation of a kink and unwinding of the alpha helix at that position. This effect could be further propagated at position 41 by the presence of a glycine residue, leading to the loss of the helical conformation of helix-2. The C-terminal domain, which includes the divergent position 96, is

regarded to play an important role in protein stability [40]. Furthermore, position S96 is often phosphorylated and involved in inducing apoptosis, mitochondrial depolarization, nuclear localization, and transduction of cells by Vpr. While not a statistically significant change with regard to Jensen-Shannon divergence, position 77 showed high diversity in both the R5 and X4 groups, and the consensus sequences both indicated a Q versus R polymorphism at this position. In clinical studies comparing HIV-1 progressor and nonprogressor patients, an R77Q variation was correlated with slower disease progression (long-term nonprogression) [36]. In vitro studies of Vpr demonstrated that the R77Q mutation attenuates Vpr in its ability to induce apoptosis in T lymphocytes [36, 41]. Notably, while position 77 of Vpr is part of the third α -helix, the region spanning positions 74–78 exhibits a less strictly defined α -helix conformation [7].

In parallel studies of Vpr, we showed that four amino acids at three positions within the primary sequence of Vpr (I37, N41, S41, and A55) were associated with significant differences in patient neuropsychological status [42]. Amino acids N41 and A55 were linked to significant increases in patient neurocognitive deficits, while I37 and S41 were associated with improvements in neurocognitive abilities. The demonstration that positions 37 and 41 are specifically linked to both coreceptor usage and neuropathogenesis is interesting in light of (i) the role of the HIV-1-infected macrophage (infected by an R5 virus) as a “Trojan Horse” that carries HIV-1 into the brain [43, 44] and (ii) the importance of coreceptor usage in HIV-1-associated neuropathogenesis [45–47].

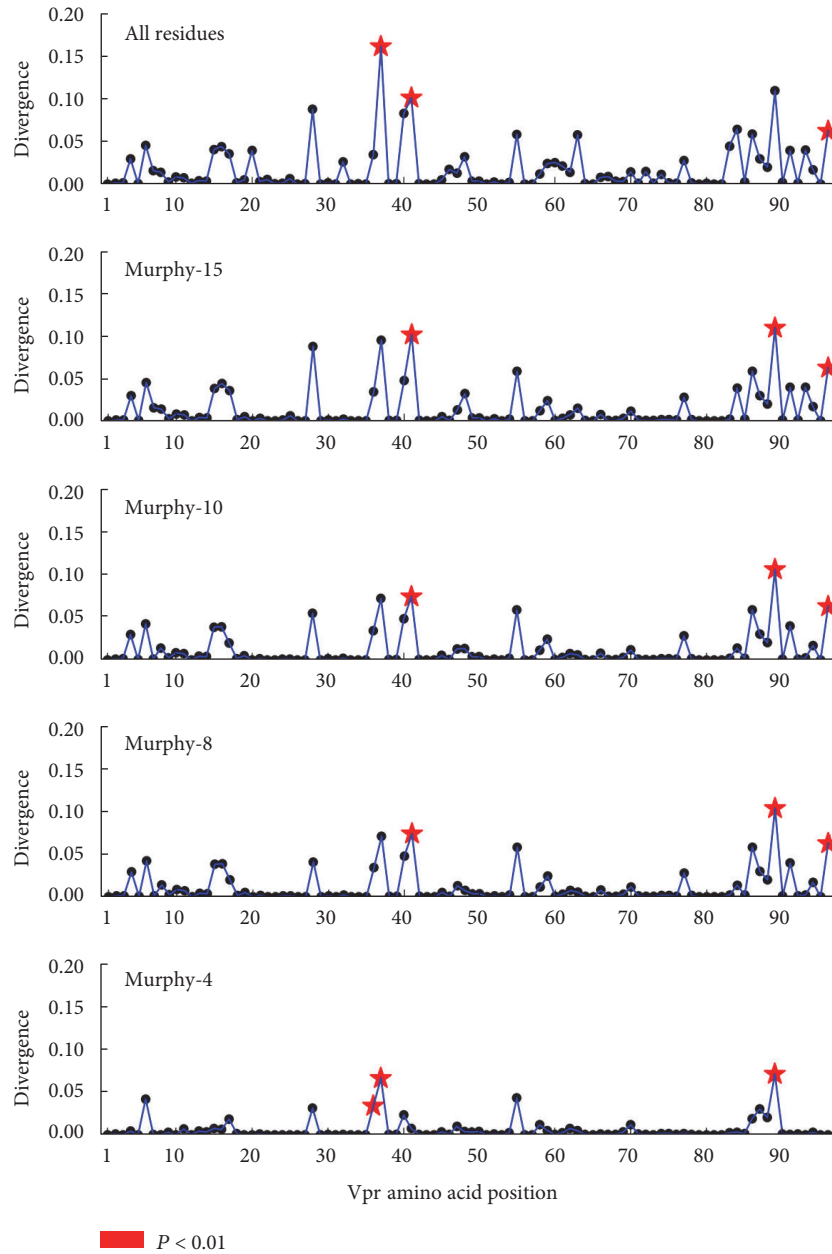


FIGURE 3: Jensen-Shannon divergence identifies positions with differential amino acid profiles between R5 and X4 Vpr sequences. The Jensen-Shannon divergence between R5 and X4 Vpr sequences was computed for each amino acid position and plotted with a black circle. Statistically divergent positions, as determined by a one-tailed Monte Carlo permutation test (1000 iterations), are indicated by a red star ($P < 0.05$). This methodology was repeated for each of the reduced alphabets, Murphy-15, Murphy-10, Murphy-8, and Murphy-4.

These results suggest that coordinated evolution takes place between Vpr and genetic determinants of viral coreceptor usage. These findings potentially add new dimensions to the already complex and multifaceted involvement of Vpr in HIV-1 infection and associated disease. In HIV-1-infected T lymphocytes and macrophages, coselection of Vpr with coreceptor phenotype may contribute to immunopathogenesis associated with productive infection, as well as the establishment and maintenance of long-lived viral reservoirs even in the face of highly effective antiretroviral therapies. In the brain, coevolution of Vpr with the

predominant R5 phenotype of viruses that infect the brain may better position Vpr as a contributor to HIV-1-associated neuropathogenesis. Finally, these studies, as do our previous studies of Tat and LTR coevolution with gp120, suggest the distinct possibility that other viral proteins also evolve in conjunction with viral coreceptor usage.

Disclosure

The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of the

TABLE 3: HIV-1 Vpr positions with high Jensen-Shannon divergence between R5- and X4-associated sequences. The calculated Jensen-Shannon divergence, along with the associated *P* value in parentheses (in italics when *P* < 0.01), is displayed for positions 36, 37, 41, 89, and 96.

Position	All residues	Murphy-15	Murphy-10	Murphy-8	Murphy-4
36	0.0346 (0.0367)	0.0341 (0.0205)	0.0341 (0.0143)	0.0341 (0.0271)	0.0331 (0.0068)
37	0.1612 (0.0053)	0.0950 (0.0322)	0.0715 (0.0555)	0.0709 (0.0389)	0.0656 (0.0026)
41	0.1012 (0.0013)	0.1012 (0.0005)	0.0735 (0.0024)	0.0735 (0.0034)	0.0072 (0.4613)
89	0.1094 (0.0103)	0.1094 (0.0029)	0.1048 (0.0027)	0.1034 (0.0013)	0.0704 (0.0034)
96	0.0623 (0.0051)	0.0623 (0.0029)	0.0623 (0.0030)	0.0623 (0.0030)	0.0000 (0.5005)

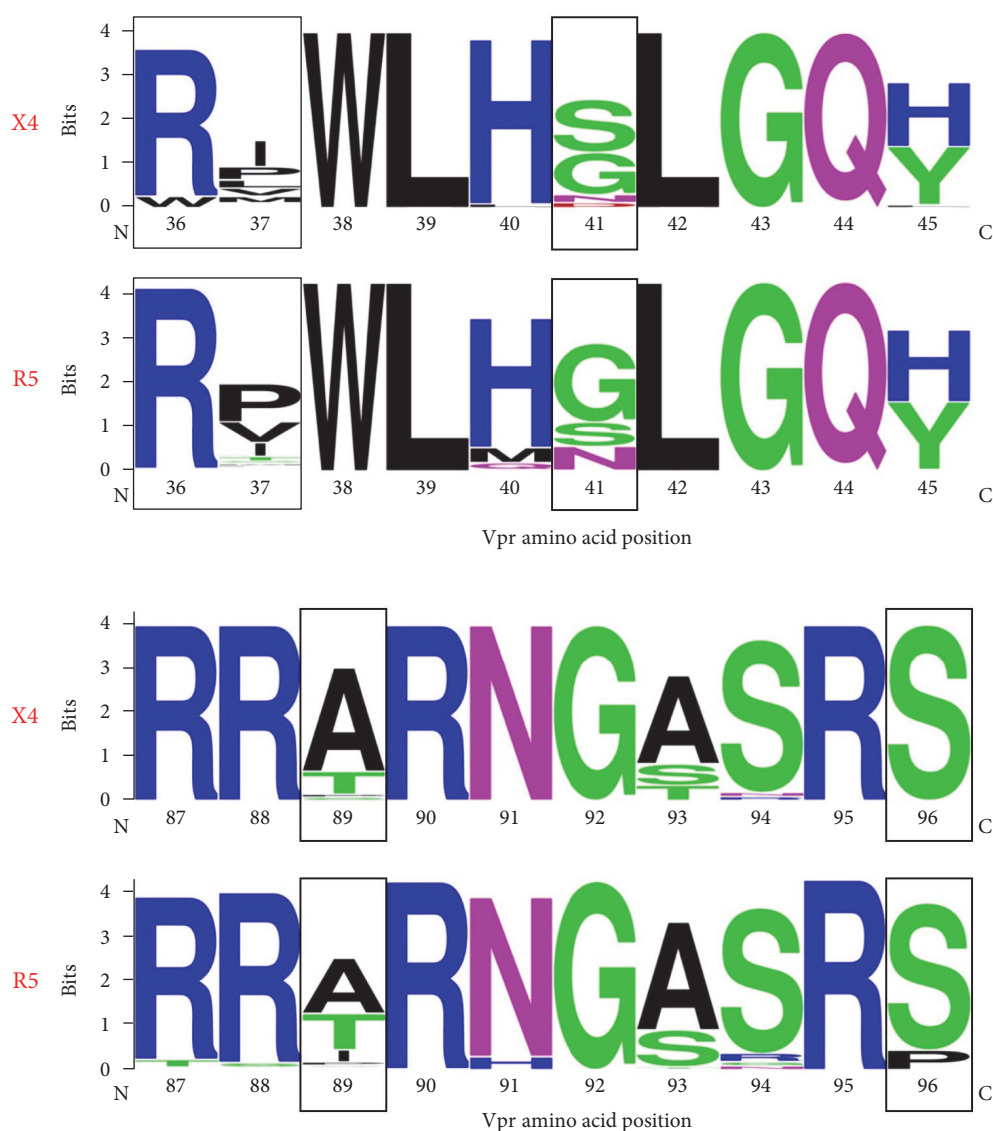


FIGURE 4: Sequence logos for Vpr regions 36–45 and 87–96. Both regions within the Vpr amino acid sequence included several peaks in amino acid diversity across multiple alphabets. They also contained all positions (36, 37, 41, 89, and 96) found to be statistically divergent using Jensen-Shannon divergence (boxed sequence logos).

data; in the writing of the manuscript; and in the decision to publish the results. The contents of the paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Gregory C. Antell, Will Dampier, Michael R. Nonnemacher, Brian Wigdahl, and Fred C. Krebs conceived and designed the study. Gregory C. Antell and Will Dampier designed and performed the bioinformatic and statistical analyses. Brian Wigdahl, Jeffrey M. Jacobson, Zsofia Szep, Michael R. Nonnemacher, Vanessa Pirrone, Katherine Kercher, Wen Zhong, Benjamas Aiamkitsumrit, Yucheng Liu, Gregory C. Antell, Tony James, Jean Williams, and Shendra Passic managed the Drexel Medicine CARES Cohort and/or obtained sequences from patient PBMC samples. Gregory C. Antell and Fred C. Krebs prepared and designed the figures and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank all patients who are part of the Drexel Medicine CARES Cohort. The authors would also like to thank the clinical staff within the Division of Infectious Diseases and HIV Medicine and Center for Clinical and Translational Medicine in the Institute for Molecular Medicine and Infectious Disease at the Drexel University College of Medicine who are involved in the recruitment, enrollment, obtaining consent, obtaining clinical histories, venipuncture, and delivery of peripheral blood to the research laboratories in the Center for Molecular Virology and Translational Neuroscience in the Institute for Molecular Medicine and Infectious Disease. These studies were funded in part by the Public Health Service, National Institutes of Health, through grants from the National Institute of Neurological Disorders and Stroke, NS32092 and NS46263, the National Institute of Drug Abuse, DA19807 (Dr. Brian Wigdahl, principal investigator), NS089435 (Dr. Michael Nonnemacher, principal investigator), a developmental grant from the National Institute of Mental Health Comprehensive NeuroAIDS Core Center (CNAC) Developmental Core, P30 MH092177 (Dr. Fred Krebs, principal investigator of the developmental grant), and under the Ruth L. Kirschstein National Research Service Award T32 MH079785 (Dr. Brian Wigdahl, principal investigator of the Drexel University College of Medicine component). Drs. Michael Nonnemacher, Will Dampier, and Fred Krebs were also supported by faculty development funds provided by the Department of Microbiology and Immunology and the Institute for Molecular Medicine and Infectious Disease.

References

- [1] K. T. Arrildt, S. B. Joseph, and R. Swanstrom, "The HIV-1 env protein: a coat of many colors," *Current HIV/AIDS Reports*, vol. 9, no. 1, pp. 52–63, 2012.
- [2] P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson, "Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody," *Nature*, vol. 393, no. 6686, pp. 648–659, 1998.
- [3] G. C. Antell, W. Dampier, B. Aiamkitsumrit et al., "Utilization of HIV-1 envelope V3 to identify X4- and R5-specific Tat and LTR sequence signatures," *Retrovirology*, vol. 13, no. 1, p. 32, 2016.
- [4] R. Swanstrom and J. Coffin, "HIV-1 pathogenesis: the virus," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 12, article a007443, 2012.
- [5] F. Gonzalez-Scarano and J. Martin-Garcia, "The neuropathogenesis of AIDS," *Nature Reviews. Immunology*, vol. 5, no. 1, pp. 69–81, 2005.
- [6] B. Aiamkitsumrit, W. Dampier, G. Antell et al., "Bioinformatic analysis of HIV-1 entry and pathogenesis," *Current HIV Research*, vol. 12, no. 2, pp. 132–161, 2014.
- [7] W. Schuler, K. Wecker, H. de Rocquigny, Y. Baudat, J. Sire, and B. P. Roques, "NMR structure of the (52-96) C-terminal domain of the HIV-1 regulatory protein Vpr: molecular insights into its biological functions," *Journal of Molecular Biology*, vol. 285, no. 5, pp. 2105–2117, 1999.
- [8] K. Wecker and B. P. Roques, "NMR structure of the (1-51) N-terminal domain of the HIV-1 regulatory protein Vpr," *European Journal of Biochemistry*, vol. 266, no. 2, pp. 359–369, 1999.
- [9] K. Wecker, N. Morellet, S. Bouaziz, and B. P. Roques, "NMR structure of the HIV-1 regulatory protein Vpr in H₂O/trifluoroethanol. Comparison with the Vpr N-terminal (1-51) and C-terminal (52-96) domains," *European Journal of Biochemistry*, vol. 269, no. 15, pp. 3779–3788, 2002.
- [10] M. Kogan and J. Rappaport, "HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention," *Retrovirology*, vol. 8, p. 25, 2011.
- [11] B. Aiamkitsumrit, W. Dampier, J. Martin-Garcia et al., "Defining differential genetic signatures in CXCR4- and the CCR5-utilizing HIV-1 co-linear sequences," *PloS One*, vol. 9, no. 9, article e107389, 2014.
- [12] C. B. Sturdevant, S. B. Joseph, G. Schnell, R. W. Price, R. Swanstrom, and S. Spudich, "Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection," *PLoS Pathogens*, vol. 11, no. 3, article e1004720, 2015.
- [13] S. B. Joseph, K. T. Arrildt, C. B. Sturdevant, and R. Swanstrom, "HIV-1 target cells in the CNS," *Journal of Neurovirology*, vol. 21, no. 3, pp. 276–289, 2015.
- [14] G. M. Shaw and E. Hunter, "HIV transmission," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 11, 2012.
- [15] C. Verhofstede, M. Nijhuis, and L. Vandekerckhove, "Correlation of coreceptor usage and disease progression," *Current Opinion in HIV and AIDS*, vol. 7, no. 5, pp. 432–439, 2012.
- [16] L. Li, B. Aiamkitsumrit, V. Pirrone et al., "Development of co-selected single nucleotide polymorphisms in the viral promoter precedes the onset of human immunodeficiency virus type 1-associated neurocognitive impairment," *Journal of Neurovirology*, vol. 17, no. 1, pp. 92–109, 2011.

- [17] N. Parikh, W. Dampier, R. Feng et al., "Cocaine alters cytokine profiles in HIV-1-infected African American individuals in the DrexelMed HIV/AIDS genetic analysis cohort," *Journal of Acquired Immune Deficiency Syndromes*, vol. 66, no. 3, pp. 256–264, 2014.
- [18] M. R. Nonnemacher, V. Pirrone, R. Feng et al., "HIV-1 promoter single nucleotide polymorphisms are associated with clinical disease severity," *PLoS One*, vol. 11, no. 4, article e0150835, 2016.
- [19] M. A. Jensen, F. S. Li, A. B. van 't Wout et al., "Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences," *Journal of Virology*, vol. 77, no. 24, pp. 13376–13388, 2003.
- [20] B. D. Aimkitsumrit, W. Dampier, J. Martin-Garcia et al., "Defining differential genetic signatures in CXCR4- and CCR5-utilizing HIV-1 co-linear sequences," *PLoS One*, vol. 9, no. 9, article e107389, 2014.
- [21] L. R. Murphy, A. Wallqvist, and R. M. Levy, "Simplified amino acid alphabets for protein fold recognition and implications for folding," *Protein Engineering*, vol. 13, no. 3, pp. 149–152, 2000.
- [22] P. J. Cock, T. Antao, J. T. Chang et al., "Biopython: freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, 2009.
- [23] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.
- [24] G. W. Schwartz and U. Hershberg, "Conserved variation: identifying patterns of stability and variability in BCR and TCR V genes with different diversity and richness metrics," *Physical Biology*, vol. 10, no. 3, p. 035005, 2013.
- [25] J. Zhang and J. R. Yang, "Determinants of the rate of protein sequence evolution," *Nature Reviews. Genetics*, vol. 16, no. 7, pp. 409–420, 2015.
- [26] G. Yona and M. Levitt, "Within the twilight zone: a sensitive profile-profile comparison tool based on information theory," *Journal of Molecular Biology*, vol. 315, no. 5, pp. 1257–1275, 2002.
- [27] G. Wang and R. L. Dunbrack Jr., "Scoring profile-to-profile sequence alignments," *Protein Science*, vol. 13, no. 6, pp. 1612–1626, 2004.
- [28] S. Shah, A. Alexaki, V. Pirrone, S. Dahiya, M. R. Nonnemacher, and B. Wigdahl, "Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I," *Virology Journal*, vol. 11, no. 1, p. 92, 2014.
- [29] W. Dampier, M. R. Nonnemacher, J. Mell et al., "HIV-1 genetic variation resulting in the development of new quasispecies continues to be encountered in the peripheral blood of well-suppressed patients," *PLoS One*, vol. 11, no. 5, article e0155382, 2016.
- [30] P. K. Tungaturthi, B. E. Sawaya, S. P. Singh et al., "Role of HIV-1 Vpr in AIDS pathogenesis: relevance and implications of intravirion, intracellular and free Vpr," *Biomedicine & Pharmacotherapy*, vol. 57, no. 1, pp. 20–24, 2003.
- [31] B. Poon and I. S. Chen, "Human immunodeficiency virus type 1 (HIV-1) Vpr enhances expression from unintegrated HIV-1 DNA," *Journal of Virology*, vol. 77, no. 7, pp. 3962–3972, 2003.
- [32] L. J. Zhao and H. Zhu, "Structure and function of HIV-1 auxiliary regulatory protein Vpr: novel clues to drug design," *Current Drug Targets. Immune, Endocrine and Metabolic Disorders*, vol. 4, no. 4, pp. 265–275, 2004.
- [33] D. N. Levy, Y. Refaeli, and D. B. Weiner, "Extracellular Vpr protein increases cellular permissiveness to human immunodeficiency virus replication and reactivates virus from latency," *Journal of Virology*, vol. 69, no. 2, pp. 1243–1252, 1995.
- [34] W. C. Goh, M. E. Rogel, C. M. Kinsey et al., "HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo," *Nature Medicine*, vol. 4, no. 1, pp. 65–71, 1998.
- [35] C. A. Guenzel, C. Herate, and S. Benichou, "HIV-1 Vpr—a still "enigmatic multitasker"," *Frontiers in Microbiology*, vol. 5, no. 127, p. 18, 2014.
- [36] D. Mogni, P. Citterio, B. Menzaghi et al., "Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients," *AIDS*, vol. 20, no. 4, pp. 567–574, 2006.
- [37] F. Re and J. Luban, "HIV-1 Vpr: G2 cell cycle arrest, macrophages and nuclear transport," *Progress in Cell Cycle Research*, vol. 3, pp. 21–27, 1997.
- [38] L. J. Zhao, L. Wang, S. Mukherjee, and O. Narayan, "Biochemical mechanism of HIV-1 Vpr function. Oligomerization mediated by the N-terminal domain," *The Journal of Biological Chemistry*, vol. 269, no. 51, pp. 32131–32137, 1994.
- [39] X. Chen, Y. Zheng, H. Li et al., "The Vpr gene polymorphism of human immunodeficiency virus type 1 in China and its clinical significance," *Current HIV Research*, vol. 9, no. 5, pp. 295–299, 2011.
- [40] S. Mahalingam, M. Patel, R. G. Collman, and A. Srinivasan, "The carboxy-terminal domain is essential for stability and not for virion incorporation of HIV-1 Vpr into virus particles," *Virology*, vol. 214, no. 2, pp. 647–652, 1995.
- [41] J. J. Lum, O. J. Cohen, Z. Nie et al., "Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis," *The Journal of Clinical Investigation*, vol. 111, no. 10, pp. 1547–1554, 2003.
- [42] W. Dampier, G. C. Antell, B. Aimkitsumrit et al., "Specific amino acids in HIV-1 Vpr are significantly associated with differences in patient neurocognitive status," *Journal of Neurovirology*, vol. 23, no. 1, pp. 113–124, 2017.
- [43] S. Kramer-Hammerle, I. Rothenaigner, H. Wolff, J. E. Bell, and R. Brack-Werner, "Cells of the central nervous system as targets and reservoirs of the human immunodeficiency virus," *Virus Research*, vol. 111, no. 2, pp. 194–213, 2005.
- [44] T. Fischer-Smith and J. Rappaport, "Evolving paradigms in the pathogenesis of HIV-1-associated dementia," *Expert Reviews in Molecular Medicine*, vol. 7, no. 27, pp. 1–26, 2005.
- [45] D. Gabuzda and J. Wang, "Chemokine receptors and mechanisms of cell death in HIV neuropathogenesis," *Journal of Neurovirology*, vol. 6, Supplement 1, pp. S24–S32, 2000.
- [46] D. Gabuzda and J. Wang, "Chemokine receptors and virus entry in the central nervous system," *Journal of Neurovirology*, vol. 5, no. 6, pp. 643–658, 1999.
- [47] F. J. Vazquez-Santiago and V. Rivera-Amill, "Envelope gene evolution and HIV-1 neuropathogenesis," *Journal of Neuroinfectious Diseases*, vol. 6, Supplement 2, 2015.