ELSEVIER

Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates

Mia Coetzer [a], Tonie Cilliers [a], Li-Hua Ping [b], Ronald Swanstrom [b], Lynn Morris [a,*]

[a] *AIDS Virus Research Unit, National Institute for Communicable Diseases, Private Bag X4, Sandringham, Johannesburg, Gauteng, 2131, South Africa*
[b] *UNC Centre for AIDS Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

Abstract

CXCR4 coreceptor usage appears to occur less frequently among HIV-1 subtype C viruses. The aim of this study was to investigate the genetic determinants within the V3 region of subtype C isolates able to use CXCR4. Thirty-two subtype C isolates with known phenotypes (16 R5, 8 R5X4 and 8 X4 isolates) were assessed. A subtype C-specific V3 heteroduplex tracking assay (HTA) was used to determine sample complexity, and nucleotide sequencing analysis was used to compare characteristics associated with CCR5 and CXCR4-using isolates. There were sufficient genetic differences to discriminate between R5 viruses and those able to use CXCR4. In general, R5 isolates had an HTA mobility ratio >0.9 whereas CXCR4-using isolates were usually <0.9. Multiple bands were more frequently seen among the dualtropic isolates. Sequence analysis of the V3 region showed that CXCR4-using viruses were often associated with an increased positive amino acid charge, insertions and loss of a glycosylation site, similar to HIV-1 subtype B. In contrast, where subtype B consensus V3 has a GPGR crown motif irrespective of coreceptor usage, all 16 subtype C R5 viruses had a conserved GPGQ sequence at the tip of the loop, while 12 of the 16 (75%) CXCR4-using viruses had substitutions in this motif, most commonly arginine (R). These findings were confirmed using a larger published data set. We therefore suggest that changes within the crown motif of subtype C viruses might be an additional pathway to utilise CXCR4 and thus GPGQ may limit the potential for the development of X4 viruses.
© 2006 Elsevier Inc. All rights reserved.

Keywords: HIV-1 subtype C; CXCR4 usage; V3 region; V3-HTA

Introduction

The binding of HIV-1 to CD4 and a coreceptor facilitates the entry of the virus into host target cells. The two major coreceptors involved are CCR5 and CXCR4 (Deng et al., 1996; Dragic et al., 1996; Feng et al., 1996), with CCR5 being the most commonly used during transmission and early infection (Michael et al., 1997). HIV-1 isolates can be differentiated based on their ability to use these coreceptors with R5 viruses (formerly known as non-syncytium-inducing or NSI viruses) utilising CCR5, X4 viruses using CXCR4 and R5X4 viruses able to use both receptors (collectively known as syncytium-inducing or SI isolates) (Berger et al., 1998). Viruses able to use CXCR4 emerge later in the course of infection in over 50% of individuals infected with subtype B (Connor et al., 1997;

Richman and Bozzette, 1994; Scarlatti et al., 1997). Subtype C is the most prevalent subtype, but to date only limited numbers of CXCR4-using viruses have been described and characterised, with R5 viruses dominating at all stages of disease, including late stage AIDS (Abebe et al., 1999; Batra et al., 2000; Bjorndal et al., 1999; Cilliers et al., 2003; Johnston et al., 2003; Ping et al., 1999). This suggests that there may be factors limiting the development of X4 viruses in subtype C, but it is not clear whether this is due to host immune or virological constraints. Determining what defines a subtype C X4 virus might increase the likelihood of identifying these viruses, improve understanding of the factors involved in their development, as well as the role CXCR4 usage plays in disease progression of subtype C HIV-1.

The third variable loop (V3) of gp120 is generally 35 amino acids long, highly variable and a critical determinant of coreceptor usage, although other regions such as the V1/V2 and the conserved region 4 (C4) have also been implicated

---

(Hartley et al., 2005; Pastore et al., 2006). Within the V3 region variation of specific amino acids, increased positive amino acid charge and glycosylation are some of the characteristics that have been identified as influencing coreceptor usage in subtype B. In particular, the amino acids at positions 11 and 25 are used to distinguish between NSI (R5) and SI (R5X4 and X4) viruses, with the latter often having a positive amino acid at these positions (De Jong et al., 1992; Fouchier et al., 1995; Hoffman et al., 2002). As a result, the net V3 charge is often a good indicator and determinant of viral tropism, with a high positive charge correlating with CXCR4 usage. An important determinant of CCR5 usage appears to be the predicted N-linked glycosylation site at position 6–8 within the V3 region as loss of this glycan is associated with less efficient usage of CCR5 and in some cases enhanced ability to use CXCR4 for entry (Ogert et al., 2001; Polzer et al., 2002). The role of these V3 genetic characteristics and the influence they may have on subtype C coreceptor usage remain to be evaluated as to date only a few isolates able to use CXCR4 have been identified in this subtype.

The V3-specific heteroduplex tracking assay (V3-HTA) has been used to investigate the presence of multiple variants within the V3 region in subtype B (Nelson et al., 1997). This assay is based on the formation of heteroduplexes between the isolate V3 sequence and a consensus V3 sequence (probe). Divergence from the consensus sequence is then measured by the mobility ratio of the heteroduplexes formed and represents the variation within the V3 region, with CXCR4-using viruses usually the most divergent (Nelson et al., 1997, 2000). A subtype C-specific V3-HTA has previously been applied to subtype C samples, but no CXCR4-using viruses were identified due to the low frequency of X4 viruses in this subtype (Ping et al., 1999). In this study, we selected 32 subtype C isolates from different patients with experimentally determined coreceptor usage profiles (16 R5 and 16 R5X4 or X4) and used the V3-HTA to screen for population complexity

of these biological variants. Furthermore, the sequences of these isolates were studied to determine if there were any distinct changes in the V3 region that might be used to differentiate CCR5 and CXCR4 usage in subtype C.

Results

V3-HTA can distinguish sample complexity within subtype C isolates with different biological phenotypes

To investigate the complexity within the V3 region of HIV-1 subtype C samples with different phenotypes, a V3-HTA was used to screen 16 R5 isolates, 8 R5X4 and 8 X4 isolates from different patients at various stages of disease, including late stage AIDS (Table 1). Two isolates from a single patient (Du179) that were 1 year apart were both included as they showed different coreceptor profiles. The V3 region from all isolates was amplified and hybridised to a subtype C R5 radiolabelled probe and resolved on a polyacrylamide gel. In general, the heteroduplexes that formed between the probe and the R5 viruses migrated faster through the gel, close to the probe homoduplex, while the R5X4 and X4 viruses migrated more slowly through the gel at variable distances and often closer to the single-stranded probe (Fig. 1). At the extreme, one R5X4 (SW30) and one X4 (TM2) had heteroduplexes that migrated above the single-stranded probe, suggesting a high degree of genetic difference and/or an insertion or deletion relative to the probe. Apart from one R5 isolate (SW4) and one X4 isolate (SW7), all the R5 and X4 isolates had single bands indicative of homogeneous populations within the V3 region. In contrast, 5 of the 8 R5X4 isolates had multiple variants often migrating at different distances between the single strand and homoduplex probe. These multiple variants suggest heterogeneous mixtures of viruses with different phenotypes and were therefore further referred to as dual/mixed tropic (DM) viruses.

Table 1
Clinical information of the 16 R5, 8 R5X4 and 8 X4 HIV-1 subtype C isolates used in this study

| R5 isolates					R5X4/X4 isolates				
Isolate	Clinical status	CD4 count (cells/μl)	Viral load (copies/ml)	Biotype	Isolate	Clinical status	CD4 count (cells/μl)	Viral load (copies/ml)	Biotype
Du151(11-98)	Acute	367	>500,000	R5	CM9	AIDS	24	NA	R5X4
PCP1	AIDS	2	NA	R5	Du179(05-99)	Chronic	279	2640	R5X4
CM1	AIDS	43	146,514	R5	Du36	Acute	25	54,944	R5X4
CM4	AIDS	47	163,755	R5	RP1	Rapid progressor	7	178,830	R5X4
SW2	AIDS	84	157,150	R5	SW20	AIDS	2	43,595	R5X4
SW3	AIDS	53	261,880	R5	SW30	AIDS	2	73,860	R5X4
SW4	AIDS	76	1,496,620	R5	TM1	Slow progressor	NA	190,000	R5X4
SW5	AIDS	40	1,374,235	R5	TM18b	Slow progressor	202	500,000	R5X4
SW8	AIDS	67	1,198,880	R5	DR28	AIDS, on treatment	173	269,000	X4
SW9	AIDS	65	301,605	R5	Du179(05-00)	Chronic	231	2228	X4
TM3	Slow progressor	329	11,178	R5	Du55	Acute	13	6589	X4
TM4	Slow progressor	692	25,815	R5	SW12	AIDS	27	68,410	X4
TM5	Slow progressor	1378	22,488	R5	SW7	AIDS	10	NA	X4
TM6	Slow progressor	846	108,716	R5	TM2	Slow progressor	NA	25,156	X4
TM10	Slow progressor	1158	685	R5	TM46b	Slow progressor	4	28,613	X4
TM12	Slow progressor	976	21,976	R5	TM9	Slow progressor	11	296,865	X4

NA: not available.
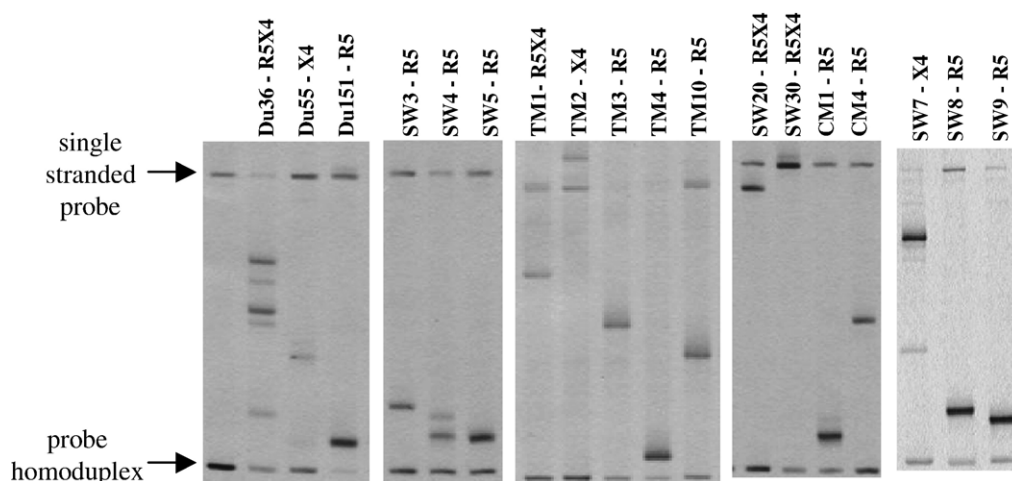
Fig. 1. V3-HTA of HIV-1 subtype C isolates with different biological phenotypes. PCR products from viral isolates were hybridised to a radiolabelled subtype C R5 probe and separated on a polyacrylamide gel. Heteroduplexes between the isolate and probe usually migrated between the single-stranded probe and the probe homoduplex as shown in five separate gels.

The mobility ratio (k) was calculated for each isolate based on the migration rate of the heteroduplex bands divided by the migration distance of the homoduplex band from the probe. In isolates with multiple bands, the mobility ratio for each variant was calculated. The majority of R5 isolates clustered closer together with a median mobility ratio of 0.94 (range 0.76–0.96) (Fig. 2A). The three R5 isolates (TM3, CM4 and TM10) with low mobility ratios ($k < 0.90$) had a single deletion within the V3 region at amino acid positions 23, 24 and 25 respectively compared to the probe sequence, which resulted in slower migration through the gel. The mobility ratios of R5X4 and X4 isolates showed a broader range (0.52–0.93) with a median of 0.64. There was no significant difference in median k values between R5X4 and X4 isolates, although almost all fell below 0.90. The one isolate with a mobility ratio above 0.90 (DR28) was from a drug treated patient and had 35 amino acids similar to the probe. A graphical display of these data relative to coreceptor usage is shown in Fig. 2B. A mobilty ratio of 0.90 separated most R5 isolates from the CXCR4-using viruses ($p < 0.001$, Fisher's Exact Test).

Analysis of isolates with multiple variants

Seven isolates in this study had multiple variants in the V3 region that were clearly distinguishable on a V3-HTA. This included one R5 isolate (SW4), five DM tropic isolates (CM9, Du36, RP1, SW30, TM1) and one X4 isolate (SW7). In order to analyse the different variants, molecular clones were made from five of these isolates (TM1 and SW7 were not cloned). The V3 amino acid alignments for each of the variants were compared, and the phenotypes of individual molecular clones were predicted. Phenotypes were predicted based on the V3 amino acid charge, sequence analysis and C-PSSM algorithm (Jensen et al., 2006).

A total of five variants were cloned from the DM tropic isolate Du36 (Fig. 3A). These variants were present in different proportions, with variants 3 and 5 being the most common. The V3 amino acid charge for all 5 variants was between 6.5 and 7.5,

and 3 of the variants (clone numbers 2, 3 and 5) lacked the predicted *N*-glycosylation site indicative of CXCR4 usage. The mobility ratios of 4 variants were consistent with CXCR4 usage, however, variant 1 had a higher sequence similarity to the R5 probe which may account for the dual tropism in this isolate. Nonetheless, the C-PSSM predicted that all clones would use CXCR4.

Sequence analysis of clones from 1 R5 and 3 additional DM isolates with multiple bands were also investigated (Fig. 3B). The amino acid sequences from both clones of SW4 were identical, and this was supported by the charge and length, which were typical of R5 viruses. The mobility differences between the 2 variants of SW4 visible on the V3-HTA were thus due to synonymous nucleotide changes. Among the 3 DM isolates, amino acid differences were seen between the 2 clones of each isolate, confirming a mixture of variants. However, the C-PSSM predicted that for CM9 and SW30 both variants used CXCR4, while for RP1 one clone was predicted to use CCR5. The genetic characterisations of the CXCR4 variants differed between patients, although some commonalities in the positions were noted. These characteristics included positively charged amino acids at positions 11 and 25, changes in the crown motif and insertions. Unfortunately, biological clones from these DM isolates were not available to confirm predicted phenotypes.

V3 sequence variability of CCR5- and CXCR4-using variants

Two data sets were compiled representing CCR5- and CXCR4-using amino acid sequences. The CCR5-using data set was obtained from the 16 patients with R5 viruses. Single sequences were obtained from 10 of the 16 patients with homogenous R5X4 or X4 viruses. For two patients (SW7 and TM1), population-based sequences were used as clones were not available. In the remaining 4 patients with DM tropic viruses where more than one clonal sequence was available, we selected the sequence with the least similarity to the probe in order to increase the sensitivity of identifying genetic characteristics associated with CXCR4 usage. These V3 sequences from the 16

A

	R5 isolates				R5X4/X4 isolates		
Isolate	Mobility ratios	# amino acids	Charge	Isolate	Mobility ratios	# amino acids	Charge
Du151(11-98)	0.96	35	3.5	CM9	0.85/0.9*	35	5.5
PCP1	0.96	35	3.5	Du179(05-99)	0.78	34	4.5
CM1	0.94	35	2.5	Du36	0.68/0.71/0.76/0.77/0.91*	35, 36	6.5, 7.5
CM4	0.8	**34**	2.5	RP1	0.62/0.84*	35, 37	4.5, 7.5
SW2	0.96	35	3.5	SW20	0.58	37	5.5
SW3	0.9	35	3.5	SW30	0.54/0.93*	35,37	7.5, 8
SW4	0.92/0.94	35	4*	TM1	0.61/0.68	37	7.5
SW5	0.95	35	3.5	TM18b	0.88	35	6
SW8	0.93	35	3.5	DR28	0.91	35	7.5
SW9	0.94	35	3.5	Du179(05-00)	0.61	32	5
TM3	0.76	**34**	3	Du55	0.83	34	6
TM4	0.96	35	4.5	SW12	0.61	37	7.5
TM5	0.91	35	4	SW7	0.65/0.78	36	7.5
TM6	0.95	35	3.5	TM2	0.52	37	5.5
TM10	0.81	**34**	4.5	TM46b	0.72	34	6
TM12	0.95	35	2	TM9	0.54	37	7
Median	0.94	35	3.5	Median	0.64	35	6

\# Number

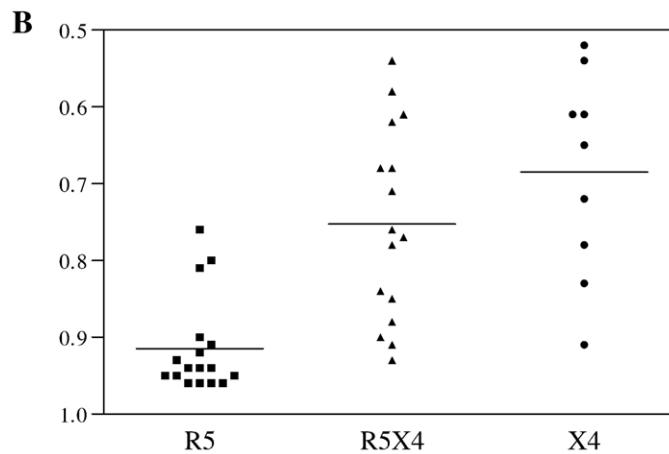* see Figure 3 for further analysis of multiple bands

B

Fig. 2. Comparison of mobility ratio's (k) and coreceptor usage of subtype C isolates. (A) V3-HTA mobility ratio, number of amino acids and charge in V3 of the variants present within the R5, R5X4 and X4 isolates. In isolates with multiple bands, all the mobility ratios are listed. (B) Association between mobility ratio (k) and biotype. A mobility ratio of 0.90 separated most R5 from R5X4 and X4 viruses.

CCR5- and 16 CXCR4-using variants were compared for length and charge. The majority of CCR5-using sequences had 34–35 amino acids with no insertions, whereas CXCR4-using sequences consisted of 32–37 amino acids due to insertions and deletions (Fig. 4A). The V3 net amino acid charge between CCR5 and CXCR4 usage was distinct with little overlap (Fig. 4B). The CCR5-using viruses had an amino acid charge between +2 and +4.5, with the majority of samples having +3.5. CXCR4-using variants ranged between +4.5 and +8 with the highest frequency of variants having a charge of +7.5. Thus, the CXCR4-using sequences had variable V3 length and increased number of positively charged amino acids.

Entropy plots were performed to compare the amount of sequence variability between the CCR5- and CXCR4-using variants (Fig. 5A). Entropy plots graphically measure the amount of variability at a specific site of an alignment, with higher entropy indicative of more variation at a specific site. In general, the CXCR4 variants were more variable across the V3 region at each specific site, with high variation at positions 11, 12, 24 and 25. The predicted N-glycosylation site (at positions 6–8) within CCR5 using viruses was conserved in all 16 R5 isolates but highly variable in the CXCR4-using samples. Similarly, the crown motif within the V3 for all R5 isolates was GPGQ, compared to CXCR4-using variants that showed variation in this motif (Fisher's Exact Test, $p < 0.0001$).

The frequency and type of amino acid substitutions at each site were determined for the CCR5- and CXCR4-using isolates (Fig. 5B). Although these results represent small numbers (16 CCR5- and 16 CXCR4-using isolates), CCR5 usage was associated with less amino acid substitutions compared to CXCR4 usage. In particular, a variety of amino acids were seen at positions 11 and 25 within CXCR4-using viruses, but these
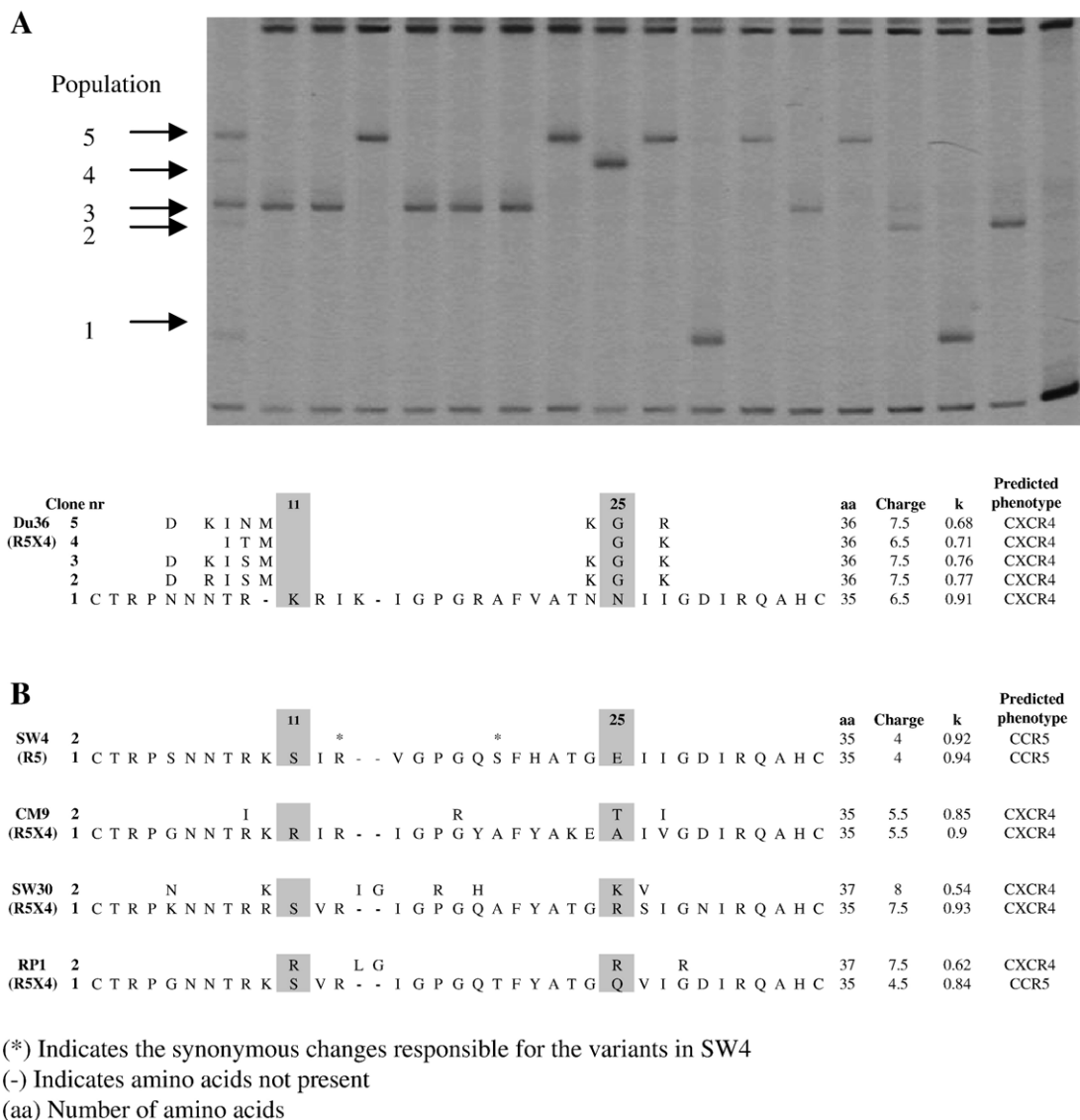
Fig. 3. Analysis of molecular clones from HIV-1 subtype C isolates with multiple variants. The variants were labelled from the bottom (highest mobility ratio) to the top of the gel. Amino acid alignments indicating only the differences within these populations, number of amino acids and charge of the V3 region are shown. Predicted phenotype based on charge, sequence analysis and C-PSSM score. (−) indicates amino acids not present, (aa) number of amino acids in the V3 region and positions 11 and 25 are highlighted. (A) Amino acid alignment of 5 populations present in Du36, as seen in accompanying gel. (B) Amino acid alignments of variants present in 1 R5 and 3 DM isolates.

were not necessarily positively charged. Twelve of the 16 CXCR4-using viruses (75%) had changes in the crown motif (GPGQ). These changes were mostly at position 18, but amino acid substitutions were also seen at positions 16 and 17.

Determining subtype C V3 characteristics associated with CXCR4 usage

Consensus sequences were compiled for the 16 CCR5 and 16 CXCR4 sequences investigated in this study. There was sequence homology for 26 of the 35 amino acids between the CCR5- and CXCR4-using isolates (Fig. 6A). Amino acid substitutions associated with the CXCR4-using data set were seen at position 5 with N being replaced by G and position 19 associated with either A or T. The consensus sequence of

CXCR4-using viruses had a 2 amino acid insertion between positions 13 and 14, as well as a further seven variable (X) amino acid positions (11, 12, 18, 23, 24, 25 and 34) compared to the CCR5 consensus. The crown motif of the CXCR4 consensus changed from GPGQ to GPGX, where X was usually an R, Y, K or H amino acid. Further analysis of the crown revealed additional genetic differences between CCR5 and CXCR4-using isolates that were of particular interest as such differences have not previously been fully explored.

In order to extend this analysis, we combined these data with previously published subtype C sequences from the Los Alamos database. This included 91 single patient sequences with determined CCR5 coreceptor usage and 17 sequences annotated as SI or CXCR4-using. This reflects the limited numbers of CXCR4-using subtype C isolates available for analysis, and the
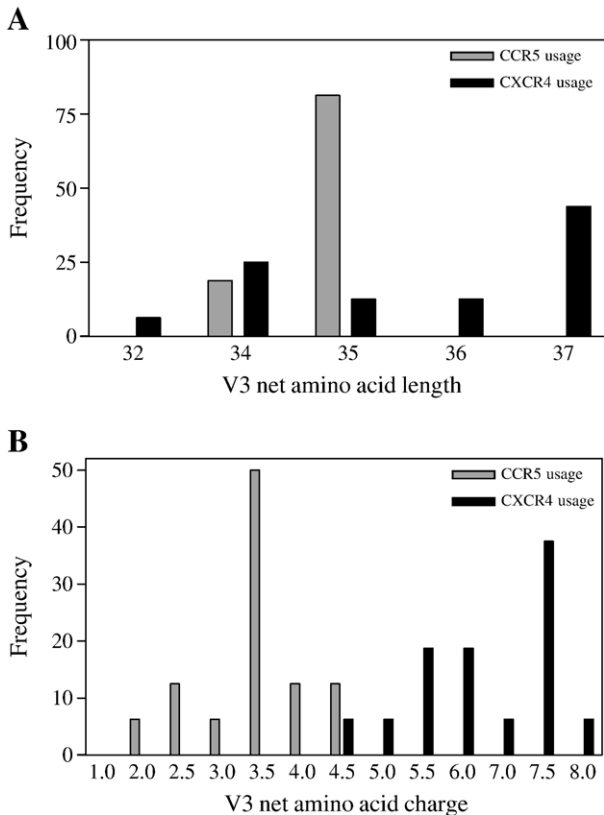
A

B

Fig. 4. Comparison of (A) V3 amino acid length and (B) V3 amino acid charge of HIV-1 subtype C isolates able to use CCR5 and CXCR4.

addition of our data set significantly boosted these numbers to 33. Since CXCR4 usage among HIV-1 subtype B has been more extensively studied, we performed a comparison with HIV-1 subtype B V3 sequences (also obtained from Los Alamos database). Using these larger data sets, we determined which genetic changes were associated with CXCR4 usage. For both subtypes B and C, there were significant associations with variable lengths, increase in charge and loss of the glycosylation site within the V3 region of CXCR4-using sequences (using contingency tables and Fisher's Exact Test, $p < 0.001$). However, changes in the GPGQ crown motif were only significant for subtype C CXCR4-using viruses ($p < 0.001$) but not for GPGR in subtype B ($p = 0.4$). Analyses showed that, while there was some variation at positions 16 (P) and 18 (R) in subtype B, this did not differ significantly between isolates that used CCR5 and those that used CXCR4 (Fig. 6B). This is in contrast to HIV-1 subtype C CXCR4-using isolates which showed marked variation particularly at position 18 from the subtype C GPGQ consensus. The entropy difference was also compared between the two data sets within each subtype using a web-based tool (ENTROPY-2 from Los Alamos data base) (Korber et al., 1994). This tool allows for a Monte Carlo randomisation within the data set, thereby providing statistical confidence of the variability associated within each data set. The analysis confirmed that there were significant variabilities ($p < 0.01$) within the crown motif between the CCR5 and CXCR4 data set within HIV-1 subtype C, whereas this was not seen in the

subtype B data sets (data not shown). Thus, there appears to be an appreciable difference between HIV-1 subtype C and B crown motif sequences and coreceptor usage.

Discussion

A subtype C-specific V3-HTA was used to examine 32 subtype C isolates with known biological phenotypes (16 R5 and 16 R5X4 or X4 isolates). Results indicated that there were sufficient genetic differences to discriminate between R5 viruses and those phenotyped as R5X4 and X4. Sequence analysis of the V3 region showed that CXCR4-using viruses were often associated with an increased number of positively charged amino acids, loss of a potential glycosylation site, as well as variable lengths. Compared to HIV-1 subtype B V3 sequences, where the consensus sequence at the GPGR crown did not differ between CCR5 and CXCR4-using isolates, the GPGQ subtype C consensus was heavily substituted in CXCR4-using viruses.

The V3-HTA has proven to be a rapid screening method to detect V3 evolutionary variants of HIV-1 subtype B and C viruses (Nelson et al., 1997, 2000; Ping et al., 1999). This assay measures distinct genetic features such as insertions, deletions or clustered amino acid changes that influence the mobility of the V3 heteroduplex. These characteristics are frequently associated with the X4-like phenotype, and using this assay, it has been possible to screen for X4 variants (Nelson et al., 1997). In a previous study of subtype C by Ping et al. (1999), no X4-like subtype C isolates were identified using a V3-HTA. This is because this study did not focus on subjects with low CD4 cell counts where X4 variants are more likely to be found, even in subtype C where such variants are rare. Here we selected subtype C viruses with experimentally determined phenotypes including a large collection of X4 variants. Using a subtype C V3-HTA, most R5 isolates had homogeneous populations with mobility ratios above 0.90. This included R5 isolates from patients with advanced disease where more heterogeneous sequences might be expected (McNearney et al., 1992). The highly conserved nature of subtype C V3 R5 isolates was also noted in the study by Ping and others (1999). The R5X4 and X4 isolates had a broader range of mobility ratios that were generally lower than 0.90. The X4 isolates had homogenous populations in the V3-HTA, whereas the DM isolates were associated with multiple variants as evidenced by multiple bands. Further analysis of these DM isolates indicated that they comprised of mixtures of variants.

A limiting factor within this study was the biased selection of samples as the ratio of R5 and X4 variants selected in this study is not reflective of prevalence of these viruses within the general populations of HIV-1 subtype C. The low numbers sampled in each class might influence the sensitivity and specificity of this assay and could therefore either improve or decline as better sampling is achieved. Among 231 HIV-1 subtype C viral isolates in our laboratory, only 10% were found to be CXCR4-using, and this was significantly correlated with a CD4 count of <200 cells/μl ($p = 0.0021$, Fisher's Exact Test). Although V3-HTA was sensitive in detecting most of the slower migrating X4 variants, this sensitivity would decrease with larger data sets due to the
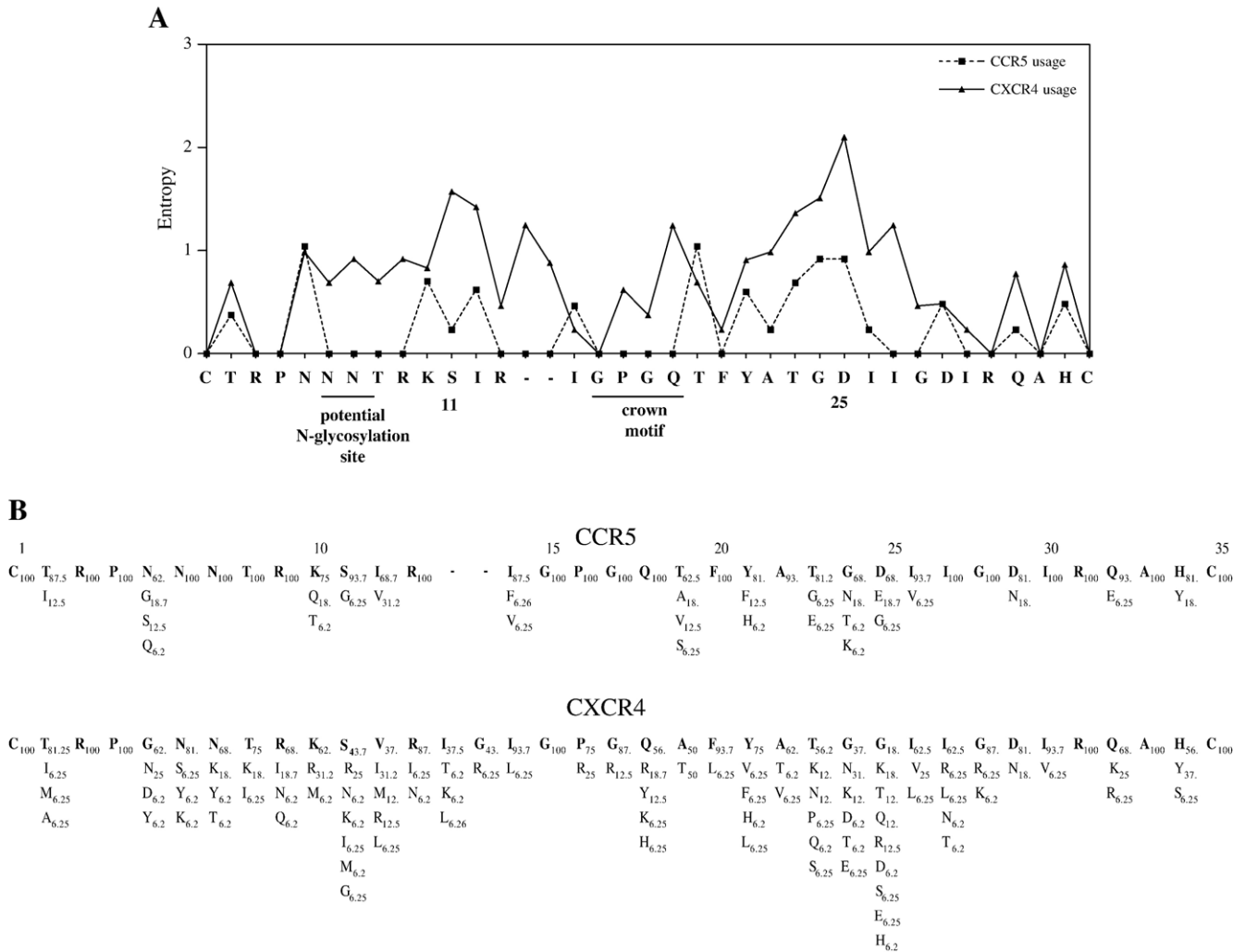
A

B

Fig. 5. Variation in the V3 region of CCR5 and CXCR4-using isolates of HIV-1 subtype C. (A) Entropy plot representing the variation at each amino acid site for CCR5 usage and CXCR4 usage. The potential *N*-glycosylation site N-[X]-T, amino acid positions 11 and 25, and crown motif are indicated. (B) Distribution of amino acids in the V3 region of 16 CCR5- and 16 CXCR4-using variants (numbers indicate frequency in total of 16 viruses).

low abundance of these variants, as well as the increase sampling of CCR5 variants with genotypic characteristics that cause slow migration (such as deletions and insertions). Similarly, the frequency of false positives (i.e. slow migrating R5 variants) would decrease the specificity. Thus, since the V3-HTA is more reflective of sample complexity, the mobility ratio criteria could vary if a larger data set was used. Despite these considerations, these data suggest that the V3-HTA was a useful screening tool to evaluate sample complexity and could assist in cloning strategies to investigate the V3 differences between CCR5- and CXCR4-using viruses.

A further caveat of this study was the selection of a single variant from the 4 DM viruses for inclusion in the data set. Since we did not know the coreceptor usage of individual clones, we selected the one with the genetic properties most commonly associated with CXCR4 usage. However, an analysis that included all the variants from the 4 DM isolates did not show significantly different outcomes (data not shown). Comparisons between the CCR5 and CXCR4 data sets revealed differences in charge and length between CCR5-

and CXCR4-using subtype C viruses. The V3 loops of CXCR4-using viruses were usually more variable and positively charged, previously shown to be associated with CXCR4 usage (De Jong et al., 1992). The increased length was usually due to one or two amino acid insertions between position 13 and 14, with amino acids I and G being the most common. Almost all CXCR4-using isolates had a high V3 charge above +4.5. This was due to the presence of increased numbers of K and R residues that were scattered throughout the V3 of CXCR4-using viruses, including positions 11 and 25 which are indicative of SI viruses in subtype B (Fouchier et al., 1992). Although the 11/25 rule is used for tropism determination, it is not clear whether basic amino acid substitutions at these sites are sufficient or necessary for CXCR4 usage (Kuiken et al., 1992; Nelson et al., 2000). In this study, these positions were not necessarily associated with positively charged amino acids in CXCR4-using viruses, although there was increased variation at these positions compared to R5 viruses. Nevertheless, similar to subtype B (Shioda et al., 1992), CXCR4 usage in subtype C was rarely due to a single
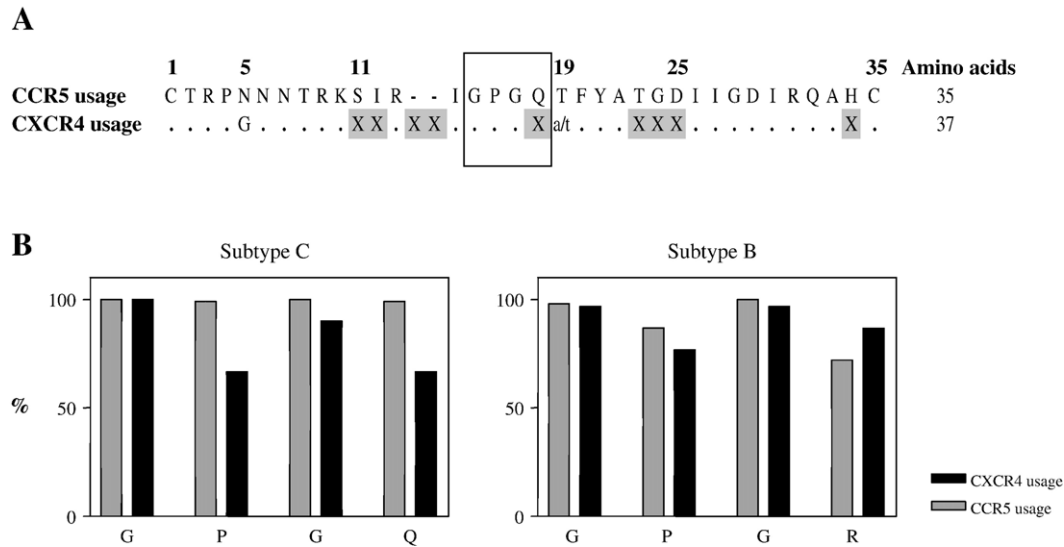
A

B

Fig. 6. (A) Consensus V3 sequences from CCR5- and CXCR4-using HIV-1 subtype C isolates. Amino acids within the CXCR4-using consensus that differ from the CCR5 consensus are highlighted and the crown motif is boxed. Variable (X) and identical (.) amino acids are indicated. (B) Comparison of amino acid variation in the V3 crown of subtype C isolates from Los Alamos and this study (107 CCR5 and 33 SI) ($p < 0.0001$) and subtype B isolates from Los Alamos (99 CCR5 and 30 CXCR4).

amino acid change but rather to changes in 3–5 amino acids that increased the length and charge of the V3 loop.

All the R5 isolates in this study had a potential *N*-glycosylation site at positions 6–8 within the V3 region. Most early viruses, able to use CCR5, have this glycan, suggesting that it is needed for CCR5 interaction (Polzer et al., 2002). As immune pressure decreases with disease progression, viruses lacking this glycan able to use CXCR4 have been shown to emerge (Pollakis et al., 2004; Polzer et al., 2002). The loss of this glycan has also been shown to assist in more efficient use of CXCR4 and thus might be an important factor in the switching of R5X4 to X4 viruses (Nabatov et al., 2004; Polzer et al., 2002). Four of the 16 CXCR4-using viruses in this study lacked this potential glycosylation site, suggesting that in subtype C this site may play a similar role. The highly conserved nature of this glycan in R5 subtype C viruses suggests it is crucial for interaction with CCR5 possibly by masking surrounding positively charged amino acids at the N-terminus of the V3 region (Hartley et al., 2005). Pollakis and others (2001) have speculated that the high frequency of V3 glycosylation within subtype C viruses might constrain the envelope structure promoting the use of CCR5 and thereby increase its transmission efficiency.

Most published subtype C isolates have a GPGQ crown motif, whereas the consensus subtype B generally contains a GPGR motif irrespective of coreceptor usage (Milich et al., 1997). In this study, CCR5-using viruses had the expected GPGQ crown, whereas this changed to GPGX, with X being R, K, H or Y in CXCR4-using isolates. Although changes within the crown of subtype C viruses have been observed by others, the possible importance of this for coreceptor usage has not been discussed (Abebe et al., 1999; Batra et al., 2000; Bjorndal et al., 1999; Cilliers et al., 2003; Johnston et al., 2003; Ping et al., 1999). Extending this analysis to sequences in the Los Alamos

database confirms the findings of a conserved crown motif in CCR5-using viruses and variable motif in CXCR4-using viruses. There were very few viruses with biologically determined CXCR4 coreceptor usage available, with most of the sequences annotated as SI phenotype (suggestive of R5X4 and X4 viruses). Nevertheless, the changes in the crown motif associated with CXCR4 usage were confirmed in the larger data set. Secondary structure prediction suggests that the GPGX forms a beta turn, and amino acid changes within this motif are critical determinants of coreceptor usage (Cormier and Dragic, 2002; Hartley et al., 2005; Hu et al., 2000; Pollakis et al., 2004; Shimizu et al., 1999; Suphaphiphat et al., 2003). Confirmation of the role of the V3 crown in coreceptor selection has been shown by recent structural studies (Huang et al., 2005; Watabe et al., 2006). In subtype B position 18 in the GPG*R*, crown was found to be less variable in X4 viruses, suggesting a functional role for R in CXCR4 usage (Resch et al., 2001), possibly predisposing subtype B viruses to using CXCR4. In subtype C, arginine (R) was the most frequent amino acid at position 18 in CXCR4-using isolates. Thus, one route to CXCR4 usage may require GPGQ to first undergo a change at position 18 to arginine (R) increasing the charge and/or altering the conformation of V3, as proposed by Hartley et al. (2005). Previous studies have shown that the transition from a R5 to X4 requires few genetic changes (at least within the V3), although these transitional intermediates may be less fit, accounting for the low frequency of X4 viruses in subtype C (Pastore et al., 2004).

In conclusion, changes within the V3 region such as increased amino acid charge, insertions, specific amino acid variation and loss of the potential glycosylation site are all factors that play a role in CXCR4 usage of subtype C viruses as have been reported in other subtypes. Although the numbers of CXCR4-using isolates in subtype C are limited, there was a noteworthy difference between the crown motif of CCR5 and

CXCR4-using viruses. We therefore suggest that there might be additional virological adaptation within subtype C viruses that allows, but also limits, these viruses to acquire the ability to use CXCR4 as a coreceptor.

Materials and methods

Isolation and coreceptor usage of HIV-1 subtype C viruses

Viral isolates were selected from previously described cohorts from our laboratory. This included isolates from adult patients with advanced HIV-1 disease [SW, CM, PCP] (Cilliers et al., 2003), sex workers with acute HIV-1 infection [Du] (Williamson et al., 2003) and one patient failing anti-retroviral treatment (Cilliers et al., 2004). Some isolates originated from a paediatric cohort of slow (TM) and rapidly progressing children (RP) (Choge et al., 2006). Levels of virus in plasma were measured using the Versant HIV-1 RNA 3.0 assay (bDNA from Bayer Nucleic Acid Diagnostics), and CD4 counts were determined using a FACS count (Becton Dickinson, San Jose, CA). Viral isolates were tested for their ability to replicate in U87.CD4 cells transfected with either CCR5 or CXCR4, as previously described (Cilliers et al., 2003; Morris et al., 2001). Biotype was assigned based on the production of p24-antigen in the CCR5 (R5 isolate) or CXCR4 (X4 isolate) cell lines. Isolates able to use both coreceptors with comparable efficiencies or within 10% of the major coreceptor were considered dualtropic (R5X4) (Berger et al., 1998).

Viral RNA isolation and RT-PCR

Viral RNA of each isolate was extracted from PBMC culture supernatant using a MagnaPure LC Isolation station and the Total Nucleic Acid isolation kit (Roche Applied Science, Penzberg, Germany). RT-PCR was performed with primers C+ V3 (5′-ATA GTA CAT CTT AAT CAA TCT GTA GAA ATT-3′) and C−V3 (5′-CCA TTT ATC TTT ACT AAT GTT ACA ATG TGC-3′), generating a 159 bp product as described (Nelson et al., 1997; Ping et al., 1999). PCR products were purified using the High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany).

V3-HTA

V3-HTA probe construction and labelling were done as previously described by Nelson et al. (1997) and Ping et al. (1999). The probe from the plasmid (D516-11) originating from a subtype C R5 virus and with only three nucleotide differences from the subtype C V3 consensus was used (Ping et al., 1999). Single-stranded probe labelling was done by digesting plasmid D516-11 with *Bam*H1 (Amersham Pharmacia Biotech, UK), end-labelling at room temperature with a mixture containing 12.5 μCi ^{35}S-dATP (Amersham Pharmacia Biotech, UK), unlabelled dGTP and Klenow DNA polymerase I (Amersham Pharmacia Biotech, UK). The probe was removed from the vector by digestion with *Spe*I (Amersham Pharmacia Biotech, UK) and purified using the High Pure PCR purification kit

(Roche Diagnostics GmbH, Mannheim, Germany) into a final volume of 50 μl. Heteroduplexes were formed between the probe and PCR product in a 10 μl reaction containing 5 μl PCR product, 3 μl labelled probe, 1 μl annealing buffer (1 M NaCl, 100 mM Tris–HCl [pH 7.5], 20 mM EDTA) and 1 μM of the C+ V3 primer denatured at 95 °C for 2 min. The reactions were cooled at room temperature for 10 min and the heteroduplexes separated in non-denaturing 12% polyacrylamide gels as described by Nelson et al. (1997). The dried gels were exposed to autoradiograms (BioMax MR, Kodak). Heteroduplex mobility ratios were determined by measuring the mobility of the slowest heteroduplexes (highest bands in the gel) of each sample and dividing it by the mobility of the probe homoduplex. The PCR products from samples with single bands were sequenced using an ABI PRISM 3100 genetic analyser with ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

Cloning

Isolates with multiple variants were selected for cloning. Purified PCR products were cloned into the pGEMTeasy vector (Promega, USA) and individual molecular clones were screened by HTA and sequenced as described above.

Analysis

All sequences were aligned with ClustalX, predicted protein translations were performed using BioEdit (version 5.0.9), phylogenetic analysis and genetic distances were determined using the MEGA program (version 2.1). The consensus sequences for isolates that used CCR5 and CXCR4 were determined using BioEdit. Additional subtype B and C V3 sequences (single sequence per patient) were downloaded from the Los Alamos database (http://www.hiv.lanl.gov). The subtype B data set contained 129 sequences with known coreceptor usage (99 CCR5 and 30 CXCR4) and the subtype C data set 91 CCR5 viruses and 17 SI viruses (since limited numbers of subtype C viruses with biologically determined CXCR4 usage were available).

Nucleotide sequence accession numbers

The sequence data for the V3 regions obtained in this study have been submitted to GenBank under the following accession numbers: DQ898249–DQ898274. The following sequences are identical to previous submissions to GenBank: AY230878, AY170658, AY529667, AY529672, AY529677, AY529676, AY529678, AY529679, AY043173, AY043174, AY529675, AF411966, DQ235616, DQ235638.

References

Abebe, A., Demissie, D., et al., 1999. HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. Aids 13 (11), 1305–1311.

Batra, M., Tien, P.C., et al., 2000. HIV type 1 envelope subtype C sequences from recent seroconverters in Zimbabwe. AIDS Res. Hum. Retroviruses 16 (10), 973–979.

Berger, E.A., Doms, R.W., et al., 1998. A new classification for HIV-1. Nature 391 (6664), 240.

Bjorndal, A., Sonnerborg, A., et al., 1999. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. AIDS Res. Hum. Retroviruses 15 (7), 647–653.

Choge, I., Cilliers, T., et al., 2006. Genotypic and phenotypic characterization of viral isolates from HIV-1 subtype C infected children with slow and rapid disease progression. AIDS Res. Hum. Retroviruses 22 (5), 458–465.

Cilliers, T., Nhlapo, J., et al., 2003. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. J. Virol. 77 (7), 4449–4456.

Cilliers, T., Patience, T., et al., 2004. Sensitivity of HIV type 1 subtype C isolates to the entry inhibitor T-20. AIDS Res. Hum. Retroviruses 20 (5), 477–482.

Connor, R.I., Sheridan, K.E., et al., 1997. Change in coreceptor use coreceptor use correlates with disease progression in HIV-1-infected individuals. J. Exp. Med. 185 (4), 621–628.

Cormier, E.G., Dragic, T., 2002. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. J. Virol. 76 (17), 8953–8957.

De Jong, J.J., De Ronde, A., et al., 1992. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J. Virol. 66 (11), 6777–6780.

Deng, H., Liu, R., et al., 1996. Identification of a major co-receptor for primary isolates of HIV-1. Nature 381 (6584), 661–666.

Dragic, T., Litwin, V., et al., 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 381 (6584), 667–673.

Feng, Y., Broder, C.C., et al., 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor [see comments]. Science 272 (5263), 872–877.

Fouchier, R.A., Groenink, M., et al., 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. 66 (5), 3183–3187.

Fouchier, R.A., Brouwer, M., et al., 1995. Simple determination of human immunodeficiency virus type 1 syncytium-inducing V3 genotype by PCR. J. Clin. Microbiol. 33 (4), 906–911.

Hartley, O., Klasse, P.J., et al., 2005. V3: HIV's switch-hitter. AIDS Res. Hum. Retroviruses 21 (2), 171–189.

Hoffman, N.G., Seillier-Moiseiwitsch, F., et al., 2002. Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop. J. Virol. 76 (8), 3852–3864.

Hu, Q., Trent, J.O., et al., 2000. Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization. J. Mol. Biol. 302 (2), 359–375.

Huang, C.C., Tang, M., et al., 2005. Structure of a V3-containing HIV-1 gp120 core. Science 310 (5750), 1025–1028.

Jensen, M.A., Coetzer, M., et al., 2006. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope v3 sequences. J. Virol. 80 (10), 4698–4704.

Johnston, E.R., Zijenah, L.S., et al., 2003. High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral treatment. J. Virol. 77 (13), 7682–7688.

Korber, B.T., Kunstman, K.J., et al., 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. J. Virol. 68 (11), 7467–7481.

Kuiken, C.L., de Jong, J.J., et al., 1992. Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype. J. Virol. 66 (9), 5704.

McNearney, T., Hornickova, Z., et al., 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. U.S.A. 89 (21), 10247–10251.

Michael, N.L., Chang, G., et al., 1997. The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. Nat. Med. 3 (3), 338–340.

Milich, L., Margolin, B.H., et al., 1997. Patterns of amino acid variability in NSI-like and SI-like V3 sequences and a linked change in the CD4-binding domain of the HIV-1 Env protein. Virology 239 (1), 108–118.

Morris, L., Cilliers, T., et al., 2001. CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis. AIDS Res. Hum. Retroviruses 17 (8), 697–701.

Nabatov, A.A., Pollakis, G., et al., 2004. Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies. J. Virol. 78 (1), 524–530.

Nelson, J.A., Fiscus, S.A., et al., 1997. Evolutionary variants of the human immunodeficiency virus type 1 V3 region characterized by using a heteroduplex tracking assay. J. Virol. 71 (11), 8750–8758.

Nelson, J.A., Baribaud, F., et al., 2000. Patterns of changes in human immunodeficiency virus type 1 V3 sequence populations late in infection. J. Virol. 74 (18), 8494–8501.

Ogert, R.A., Lee, M.K., et al., 2001. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. J. Virol. 75 (13), 5998–6006.

Pastore, C., Ramos, A., et al., 2004. Intrinsic obstacles to human immunodeficiency virus type 1 coreceptor switching. J. Virol. 78 (14), 7565–7574.

Pastore, C., Nedellec, R., et al., 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J. Virol. 80 (2), 750–758.

Ping, L.H., Nelson, J.A., et al., 1999. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. J. Virol. 73 (8), 6271–6281.

Pollakis, G., Kang, S., et al., 2001. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. J. Biol. Chem. 276 (16), 13433–13441.

Pollakis, G., Abebe, A., et al., 2004. Phenotypic and genotypic comparisons of CCR5- and CXCR4-tropic human immunodeficiency virus type 1 biological clones isolated from subtype C-infected individuals. J. Virol. 78 (6), 2841–2852.

Polzer, S., Dittmar, M.T., et al., 2002. The N-linked glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120 affects coreceptor usage, cellular tropism, and neutralization. Virology 304 (1), 70–80.

Resch, W., Hoffman, N., et al., 2001. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology 288 (1), 51–62.

Richman, D.D., Bozzette, S.A., 1994. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. J. Infect. Dis. 169 (5), 968–974.

Scarlatti, G., Tresoldi, E., et al., 1997. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. Nat. Med. 3 (11), 1259–1265.

Shimizu, N., Haraguchi, Y., et al., 1999. Changes in and discrepancies between cell tropisms and coreceptor uses of human immunodeficiency virus type 1 induced by single point mutations at the V3 tip of the env protein. Virology 259 (2), 324–333.

Shioda, T., Levy, J.A., et al., 1992. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. U.S.A. 89 (20), 9434–9438.

Suphaphiphat, P., Thitithanyanont, A., et al., 2003. Effect of amino acid substitution of the V3 and bridging sheet residues in human immunodeficiency virus type 1 subtype C gp120 on CCR5 utilization. J. Virol. 77 (6), 3832–3837.

Watabe, T., Kishino, H., et al., 2006. Fold recognition of the human immunodeficiency virus type 1 V3 loop and flexibility of its crown structure during the course of adaptation to a host. Genetics 172 (3), 1385–1396.

Williamson, C., Morris, L., et al., 2003. Characterization and selection of HIV-1 subtype C isolates for use in vaccine development. AIDS Res. Hum. Retroviruses 19 (2), 133–144.