CCR5 use by human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N-linked glycosylation site

Peter Clevestig,1 Lotta Pramanik,1 Thomas Leitner2 and Anneka Ehrnst1,3

1,3Microbiology and Tumor Biology Center1 and Department of Laboratory Medicine, Division of Clinical Virology, Karolinska University Hospital Huddinge2,3, Karolinska Institutet, Box 280, SE-171 77 Stockholm, Sweden

2Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA

Human immunodeficiency virus type 1 (HIV-1) enters cells through the chemokine receptors CCR5 (R5 virus) and/or CXCR4 (X4 virus). Loss of N-linked glycans and increased net charge of the third variable loop (V3) of the gp120 envelope glycoprotein have been observed to be important steps towards CXCR4 use. All reported sequences using CCR5 or CXCR4 exclusively, or using both, were gathered from the Los Alamos HIV Database and analysed with regard to the V3 N-linked glycosylation motifs (sequons) and charge. The V3 loop glycan had a sensitivity of 0·98 and a 0·92 positive predictive value in the context of CCR5 use. The difference from X4 was remarkable (P<0·01). Especially, the sequon motif NNT within the V3 loop was conserved in 99·2% of the major clades. The results suggest a close association between the V3 loop glycan and CCR5 use and may provide new insight into HIV-1 tropism and help to improve phenotype-prediction models.

To enter cells, human immunodeficiency virus type 1 (HIV-1) uses the coreceptors CCR5 and/or CXCR4, together with the T-cell differentiation antigen CD4. Virus using the CCR5 receptor (R5 virus; Berger et al., 1998) or of the non-syncytium-tropic phenotype is suggested to be more transmissible (Zhang et al., 1993), whilst virus using the CXCR4 receptor (X4 virus; Berger et al., 1998) has been associated with disease progression (Connor et al., 1997; Scarlatti et al., 1997).

The third variable (V3) loop of the HIV-1 gp120 envelope glycoprotein was recognized early on to play an important role in governing the choice of target cells (Hwang et al., 1991; De Jong et al., 1992; Fouchier et al., 1992; Shioda et al., 1992). Few amino acid substitutions and an increasing net charge of the V3 loop were sufficient to confer a change in cellular tropism in vitro (De Jong et al., 1992; De Wolf et al., 1994). A decreased number of N-linked glycosylation sites (sequons) in gp120, especially within and around the V3 region, has been demonstrated during evolution from the R5 to the X4 phenotype (Pollakis et al., 2001; Polzer et al., 2001, 2002).

Computer-based models have been developed to predict the biological phenotype of HIV sequences (Briggs et al., 2000; Resch et al., 2001; Jensen et al., 2003; Pillai et al., 2003), based on multiple linear regression (Briggs et al., 2000), specific amino acids within V3 and overall charge among subtype B viruses (Resch et al., 2001) and other subtypes (Pillai et al., 2003). Machine learning has been used to develop phenotype classifiers (Resch et al., 2001) and position-specific scoring matrices have also been used (Jensen et al., 2003).

These methods relate to the assumption or possibility that the R5 and X4 phenotypes may be evaluated or classified, based on the properties of their specific amino acids, from the same scale or sum of scales. If the R5- and X4-specific properties relate to different characteristics, however, such programs may overlook essential differences in the R5 versus X4 phenotypes.

In a previous study (Clevestig et al., 2005), a phylogenetic analysis was performed on a large set of individual HIV-1 sequence clones from integrated DNA or plasma RNA, connected to sequences of isolates with determined coreceptor use. We made a preliminary observation (Clevestig et al., 2005) that R5 isolates and R5-associated sequences seemed to be connected to the presence of the N-linked glycosylation motif within the V3 loop of the env gp120 gene. As we studied only four individuals, linked epidemiologically as two mother–child pairs, we sought to explore this finding in the most epidemiologically and genetically

Tables showing details of sequences retrieved and sequence alignments are available as supplementary material in JGV Online.
diverse HIV-1 strains available. We retrieved sequences from the Los Alamos Database in a fashion similar to that of Resch et al. (2001), but included all available HIV-1 types, subtypes and recombinant forms.

We retrieved 1015 sequences, covering the V3 loop of all HIV-1 groups, group M subtypes, sub-subtypes and circulating recombinant forms (CRFs), representing all submitted sequences with a reported coreceptor use as of October 2004. Among the sequences were many multiples per patient, from which we performed a random selection to choose one patient sequence so that all sequences had a unique identity code. The selection was randomized through lottery and without consideration of the different phenotypes present.

We included sequences from HIV-1, spanning at least the V3 loop of the gp120 env gene with a specified coreceptor use, such as CCR5, CXCR4 or dual-tropic (CCR5 and CXCR4). We excluded sequences without a specified unique individual code and/or sequences using other coreceptors. Two amino acid sequences were found to be identical, despite unique identity codes. One sequence was therefore omitted randomly.

We assessed the glycosylation sequon motifs in the remaining 176 HIV-1 sequences with known coreceptor use from the database in Los Alamos (http://www.hiv.lanl.gov). All sequences were aligned manually and translated by using the sequence alignment editor Se-Al (Rambaut, 2002), also available at http://evolve.zoo.ox.ac.uk/software.html?id=seal.

The sequons were governed by the amino acid order asparagine–X–threonine/serine–Y (N–X–S/T–Y) (Marshall, 1972), where X can be any amino acid except proline (P) in the threonine (T) context (Gavel & von Heijne, 1990; Kasturi et al., 1997; Mellquist et al., 1998) and also not tryptophan (W), aspartic acid (D) or glutamine (E) in a serine (S) context (Kasturi et al., 1997). For the Y position, only proline would completely hinder oligosaccharide addition through steric hindrance in both contexts (Gavel & von Heijne, 1990; Mellquist et al., 1998). These parameters provided us with a high probability of oligosaccharide addition (Gavel & von Heijne, 1990; Shakin-Eshleman et al., 1996; Kasturi et al., 1997; Mellquist et al., 1998) and the criteria for evaluating each sequon as a possible N-linked glycosylation site. Sequences lacking data or exhibiting ambiguities in a site were excluded from this calculation.

The amino acid net charge of the V3 loop for each sequence at pH 7·0 was determined by using an online peptide-property calculator from Innovagen (http://www.innovagen.se).

A two-tailed Fisher’s exact test was used to compare the significance of the presence or lack of glycosylation sequons in conjunction with CCR5, CXCR4 or dual-tropic use, and a one-tailed test in the analysis of net charge between R5 and X4, and R5 and dual-tropic strains, respectively. The level of significance was set to a probability (P) of ≤0·05.

Among the 176 retrieved sequences (see Supplementary Table 1, available in JGV Online), 133 used CCR5, 29 used CXCR4 and 14 were dual-tropic (see Supplementary Table 2, available in JGV Online). However, this does not discount the possibility of inclusion of more than one sequence per patient that may have been submitted in another context, as was in fact the case with two sequences.

The dataset included HIV-1 of groups M (major, n = 171), O (outlier, n = 4) and N (non-M, non-O, n = 1). M included subtypes A (n = 22), B (n = 62), C (n = 28), D (n = 12), G (n = 5) and one unknown (U); circulating recombinant forms CRF01_AE (n = 20), CRF02_AG (n = 11); and other recombinants AC (n = 4), AD (n = 1), A1DGI (n = 1), AU (n = 1), BC (n = 1) and BF1 (n = 2).

We investigated the presence of eight possible N-linked glycosylation sequons across ~100 aa spanning the V3 region, where the fifth site resided inside the V3 loop. The frequency of sequons at each site is shown in Fig. 1.

Whilst leucine (L) in the X position has been associated with around 40 % oligosaccharide addition (Gavel & von Heijne 1990; Shakin-Eshleman et al., 1996; Kasturi et al., 1997; Mellquist et al., 1998), this has been greatly influenced by the amino acid in the Y position (Mellquist et al., 1998). Our NLS-Y sequons, most of which occurred in the sixth site (n = 20/23), involved Y-position amino acids that had a high probability of glycosylation, altogether ≥70 %.

No significant difference between the R5 and X4 sequences was observed in seven out of eight glycosylation sites (P > 0·2 for each site), whereas the difference at the fifth site was remarkable (P ≤ 4·6 × 10−12). There was also a tendency towards significance when comparing R5 with dual-tropic sequences (P = 0·058). A glycosylation sequon (none with an NLS-Y sequon) was present in 128/133 (96 %) of R5 sequences versus 11/29 (38 %) in X4 sequences and 11/14 (79 %) in dual-tropic sequencess. Further, the NNT sequon in the fifth glycosylation site was also conserved within the R5 sequences of known M-group subtypes in all but one sequence (120/121; 99·2 %).

The visual difference between R5 and dual-tropic sequons in site 7 was not significant (P = 0·12). The same was observed between dual-tropic and X4 sequons (P = 0·14). The stable glycosylation sequons at the flanking sites suggested that these were similarly important, but without discriminating between R5 and X4 strains.

To discern the role of the V3 loop net charge in relation to phenotype, the amino acid charge was calculated (Fig. 2). As expected, the net charge of R5 viruses (median 3·0) was lower than that of X4 viruses (median 5·9) (P = 8·8 × 10−12) and of dual-tropic strains (median 5·5) (P = 0·002). A breaking point was discerned at a charge of +4·2, below which there were no X4 sequences and above which there were only 10/133 (7·5 %) R5 sequences.
In X4 sequences, the fifth sequon was present in 11/29 sequences above the breaking point, $P > 0.3$ (Table 1a). There was also no significant association between charge and the fifth sequon among the dual-tropic strains ($P > 0.4$).

We calculated the sensitivity, specificity and predictive values of the fifth glycan site for CCR5 use and net charge above $+4.2$ for CXCR4 use within the M group (Table 1b). The sensitivity was very high in both cases ($\geq 0.98$).
Moreover, the high positive predictive value of the fifth glycan and CCR5 use indicated a strong link between the two. However, the stringent selection criteria provided us with relatively few X4 sequences, which may have influenced the lower predictive values.

Our analysis of available HIV-1 sequence data has provided evidence that the fifth N-linked glycosylation site is conserved in the vast majority of the available primary HIV-1 sequences using the CCR5 coreceptor. This pronounced correlation suggested that the V3 loop sequon, and the subsequently added sugar residues, may play an important role in the interaction with the CCR5 coreceptor. In contrast, X4 viruses seemed to rely preferentially on an increased net charge with a breaking point similar to that described previously (Fouchier et al., 1992).

We do not discount the possibility of other regions on the gp120 molecule also playing a role, either in conjunction with or independently of the V3 loop, but through this study, we would like to present a different perspective on the role of glycans in determining receptor tropism. Previous studies have shown a correlation between the loss of N-linked glycans and basic amino acid substitutions with emerging X4 and dual-tropic viruses (De Jong et al., 1992; Briggs et al., 2000; Polzer et al., 2001, 2002; Pontow & Ratner, 2001; Jensen et al., 2003). Our results are in agreement with previous findings, but emphasize the importance of the fifth N-linked glycan in the context of R5 viruses of the major (M) group of HIV-1. Loss of this glycan may not be enough to acquire the X4 phenotype, where additive changes seem necessary, providing the molecule with a higher net charge (Jensen et al., 2003). Similarly, Losman et al. (1999) found that the presence or absence of the V3 loop glycan did not influence the infectivity of X4 viruses.

It cannot be deduced from the information in the Los Alamos Database whether dual-tropic strains represented clones possessing the capacity to infect both CCR5 and CXCR4 cells or a mixture of R5 and X4 virus strains. Our data may be interpreted so that truly dual-tropic viral strains would be expected to contain the fifth sequon and a positive charge of over 4.2. Sequences lacking the fifth sequon and having a low charge would instead be expected to be less infectious or to use alternative coreceptors. Similarly, in a recent review, Hartley et al. (2005) suggested that dual-tropic strains would retain their CCR5-using property along with acquiring a higher net charge.

It might, therefore, be suggested that some of the original samples, from which the dual-tropic sequences studied here were derived, could have contained a mixture of R5 and X4 viral strains, rather than a population where the dual-tropic property was present in the same strain.

The amino acid sequence does not reveal the nature of the glycosylations with regard to actual addition of glycans or their number. The actual glycosylation may further depend on proximity to the C terminus (Gavel & von Heijne, 1990; Nilsson & von Heijne, 2000) and other proximal glycans (Pollakis et al., 2001). It is also important to include analysis of glycans in the other variable regions of gp120 (Pollakis et al., 2001). However, the presented data provide evidence that the V3 loop glycan appears to be important for CCR5 use.

The present interpretation of the relative role of N-linked glycosylation and charge for the use of CCR5 and CXCR4,
respectively, points to the necessity to further elucidate the relative differences/similarities of the CCR5 and CXCR4 receptors and their interaction with their natural ligands or HIV-1 glycoproteins.

Interestingly, the coreceptors CCR5 and CXCR4 are structurally similar. There is an implication that the second extracellular loops (ECL2) of CCR5 and CXCR4 interact with their respective HIV-1 viruses (Brelot et al., 1999; Chabot & Broder, 2000; Pontow & Ratner, 2001). It is interesting to note that CXCR4 has two N-linked glycosylation sites, one in ECL2 and one in the N terminus, that are absent on CCR5 (Chabot et al., 2000). Elimination of these glycans in CXCR4 broadens its coreceptor capacity to include several R5 strains, without affecting entrance of X4 strains (Chabot & Broder, 2000). Instead, eliminating CXCR4 charge may decrease its binding of X4 strains (Brelot et al., 1999).

The binding of R5 viruses of the M group to the CCR5 coreceptor may be associated with the presence of V3 loop N-linked glycosylation and a corresponding absence in the CCR5 coreceptor. The glycan may provide R5 strains with a narrow fitness in affinity to the N terminus and/or ECL2 of CCR5, which maybe lacking in X4/CXCR4 interactions. Instead, these seem to be governed by V3 loop charge (De Jong et al., 1992) and also by the CXCR4 ECL2 charge (Brelot et al., 1999; Chabot & Broder, 2000).

The association between the V3 loop glycosylation site and CCR5-using viruses could be incorporated into current and future prediction algorithms, where presence of this glycian would increase the probability of an R5 origin, whilst lack of it would not be considered a marker for an X4 origin. It is of interest that the three R5 M-group sequences, lacking the fifth sequon, had the amino acid motifs TNT (Gao et al., 1996), HNT (Engelbrecht et al., 2001) and NNA (Tebit et al., 2002), differing by only 1 nt each from the NNT motif.

A question that arises is whether this R5 conserved sequence or other conserved glycosylation motifs could be used as new targets for prophylactic interventions, especially as R5 is suggested to be the major transmissible variant of HIV-1.

Acknowledgements
We gratefully acknowledge all who submitted sequences. Financial support was provided by the Karolinska Institutet Research Funds, and the Sven Jerring Foundation to F. C. and A. E., respectively; the Swedish Research Council [grants no. K2003-16X-14724-01A (A. E.) and no. 521-2002-6449 (T. L.)] and a DOE–NIH interagency agreement (Y1-AI-1500-04) (T. L.).

References
Cleve´ stig, P., Malskovic, I., Casper, C. & 7 other authors (2005). The X4 phenotype of HIV type 1 evolves from R5 in two children of mothers, carrying X4, and is not linked to transmission. AIDS Res Hum Retroviruses 21, 371–378.


