Charged amino acid patterns of coreceptor use in the major subtypes of human immunodeficiency virus type 1

Lotta Pramanik, Ulrik Fried, Peter Clevestig and Anneka Ehrnst

Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, PO Box 280, SE-171 77 Stockholm, Sweden

Human immunodeficiency virus type 1 has several genetic subtypes and two coreceptor use phenotypes: R5 that uses CCR5, while X4 uses CXCR4. A high amino acid charge of the envelope glycoprotein 120 V3 region, common at positions 11 and 25, is important for CXCR4 use. We characterized charged V3 amino acids, retrieving all biologically phenotyped sequences from the HIV Sequence Database. Selecting individually unique ones randomly yielded 48 subtype A, 231 B, 180 C, 37 D and 32 CRF01_AE sequences; 482 were R5 and 46 were X4. Charged amino acids were conserved in both R5 and X4 with general and subtype-specific patterns. X4 viruses gained a higher charge from positive amino acids at positions other than in R5, and through the loss of negative amino acids. Other positions than 11/25 had a greater impact on charge \( P < 0.001 \). This describes how R5 evolves into X4 in a subtype-specific context, useful for computer-based predictions and vaccine design.

To determine the biological function by the amino acid sequences, generates a need to correctly couple these sequences to the appropriate function. For the highly variable human immunodeficiency virus type 1 (HIV-1), such conformity is crucial.

Despite the genetic diversity of HIV-1 in subtypes, subtypes and recombinant forms, there are only two major phenotypes for cell entry: R5 and X4 (and dual-tropic R5X4), depending on if the virus uses the chemokine coreceptors CCR5, CXCR4 or both (Berger et al., 1998). Previously, such HIV-1 phenotypes were related to whether the cells were syncytium-inducing (SI) or non-syncytium-inducing (NSI), or expressed macrophage (M-tropic) or T-cell (T-tropic) tropism. These three overlap without being interchangeable (Berger et al., 1998).

The NSI phenotype appears early in infection and is related to transmission (Zhu et al., 1993), while the SI is associated with disease progression (Koot et al., 1993). The same has been shown for the R5 and X4 phenotypes (Connor et al., 1997; Scarlatti et al., 1997; Casper et al., 2002).

HIV-1 subtypes differ epidemiologically and clinically. Subtype C is the most common. Subtype B has dominated among HIV-1-positive intravenous drug users and men who have sex with men. Subtype D is primarily found in central Africa, where subtype A is also prevalent. Subtype D had a larger proportion of CXCR4-using viruses than subtype A (Huang et al., 2007), and was associated with a faster disease progression than A, B or C (Vasan et al., 2006; Baeten et al., 2007; Kiwanuka et al., 2010; Easterbrook et al., 2010). Circulating recombinant form (CRF) 01_AE is spreading in South-east Asia.

The HIV-1 envelope glycoprotein 120 (gp120) third variable region (V3) is important for viral tropism (Hwang et al., 1991). Substitutions of single amino acids within V3 affect the NSI/SI phenotype in subtypes A, B, C, D and CRF01_AE (de Jong et al., 1992b; Fouchier et al., 1992; De Wolf et al., 1994). Positively charged V3 amino acids have been associated with the SI phenotype (de Jong et al., 1992a) and M-tropic viruses with a lower V3 net charge (Shioda et al., 1992).

When R5 turned into X4, the sequon for the potential N-linked glycosylation site within V3 was lost (Pollakis et al., 2001; Polzer et al., 2001, 2002). Others found this sequon to be necessary for CCR5, but not for CXCR4 use (Ögert et al., 2001; Clevestig et al., 2006).

Using computer models to predict R5 and X4 phenotypes requires that the latter are distinct in general and in different subtypes. We applied stringent criteria for inclusion of R5 and X4 sequences, and characterized the charged V3 amino acids of different subtypes in relation to coreceptor use.

One sequence from each available individual was selected at random from the Los Alamos HIV Sequence Database to minimize bias, reducing the available 1885 R5 and 422 X4...
sequences to 482 R5 and 46 X4 sequences of the five major subtypes. Sequences from other subtypes were discarded as there were only a few. The material consisted of subtypes A \((n=48)\), B \((n=231)\), C \((n=180)\), D \((n=37)\) and CRF01_AE \((n=32)\). Laboratory strains were excluded. The selection process, phenotype source and GenBank accession numbers are found in Supplementary Material and Supplementary Fig. S1, available in JGV Online.

The R5 phenotype dominated in all subtypes with a range of 78–95 % (Supplementary Table S1, available in JGV Online). We found that subtype C had more R5 sequences, \(P=0.03\), similar to previous findings. CRF01_AE had proportionally more X4 sequences, \(P=0.015\).

R5 sequences can be regarded as a background reference for X4, since X4 viruses are thought to emanate from them. As charge is important for CXCR4 use, we compared X4 and R5 charged amino acids. The charge of V3 sequences was based on the fact that arginine (R) and lysine (K) have a charge of +1, histidine (H) +0.1, and aspartic acid (D) and glutamic acid (E) −1. We observed several positions with charged amino acids that were more or less conserved in R5 sequences in all of the studied subtypes (Fig. 1, shaded columns).

Three positions (3, 9 and 31) were dominated by positively charged amino acids (Fig. 1, shaded in red), with R being present in approximately 100 % of the sequences. Two more positions (10 and 34) were dominated by positively charged amino acids in most but not all of the studied subtypes. At position 10, 91 % were positively charged amino acids in subtypes A, B and C. Position 34 was dominated by H in subtypes A, B, C and D. CRF01_AE displayed tyrosine (Y) in 24/25 sequences instead.

The positively charged subtype-specific positions were 13, 18, 22, 24 and 32 (Fig. 1, red boxes). Position 13 was unique in each subtype. R is present in the all consensus sequence from the HIV database, while here R dominated only in subtype C. Subtypes A and D differed in the proportion of H and R, respectively. In subtype B, position 18 had a positively charged amino acid in 90 % of the sequences.

The more complex features of amino acid distribution in subtype D had H at position 24 in 24 % of the sequences. A similar number of sequences displayed a gap or the neutral amino acid asparagine (N) at this position. At position 10 of subtype D glutamine (Q) dominated unlike in subtypes A, B, and C, which harboured a K here. However, the remaining 30 % in subtype D were K and R.

The CRF01_AE sequences displayed two subtype-specific positively charged positions: 22 and 32. The loss of positively charged amino acids at position 10 was compensated for by a positive charge at position 32, resulting in an average charge only slightly different from the other subtypes.'
Negatively charged amino acids (D and E) dominated two positions, 25 and 29, in the R5 phenotype (Fig. 1, shaded in blue). At position 25, amino acid D was present in 71–100% of subtype A, B, C and CRF01_AE sequences, followed by amino acid E. Subtype D had no amino acid at all (gap) at this position in 64% of the sequences, but the second and third most common amino acids were E (5/33) and D (4/33), respectively. Position 29 was dominated by D in all five subtypes with 82% of the sequences being a negatively charged amino acid, followed by N.

There were also three subtype-specific negatively charged positions in the R5 sequences (Fig. 1, blue boxes). At position 25 in subtype B the frequency of E was similar to D, distinguishing it from subtypes A, C and CRF01_AE, which had mostly or only D. In subtype D, position 25 was not dominated by a negatively charged amino acid.

The subtype-specific patterns of median charge distribution in the R5 and X4 sequences varied only according to the number of H residues, resulting in approximately +3 for R5 and approximately +6 for X4. The subtype-specific cut-off below which there were no X4s varied from +3.3 (subtype B) to +5.3 (subtype A). The highest R5 charge was +7.1 (subtype A), +6.2 in D and +5.0–5.2 in the other three subtypes. Within the subtypes the difference in charge between R5 and X4 strains was statistically significant (the median test): \( P = 0.05 \) (subtype A), \( P = 1 \times 10^{-7} \) (subtype B), \( P = 8 \times 10^{-5} \) (subtype C), \( P = 0.005 \) (subtype D) and \( P = 0.0002 \) (CRF01_AE). Positive charge could be acquired in almost every position in V3 apart from positions 15, 20 and 26, and the conserved flanking amino acids (positions 1, 2, 33 and 35).

The subtype-specific and generally conserved positively charged amino acid positions that were observed in R5 sequences were also seen in X4, here displayed in bold red in Fig. 2. Nevertheless, the X4 sequences had a higher overall charge than R5. This was attributed to the acquisition of positively charged amino acids other than the subtype-specific and generally conserved ones, and the loss of negatively charged amino acids (blue boxes in Fig. 2).

The negatively charged amino acids (bold blue in Fig. 2) at positions 25 and 29, generally conserved in R5, were lost in many X4 sequences. The exchange of a negatively for a positively charged amino acid added +2 in charge. At position 25, subtypes B, C, D and CRF01_AE had exchanged D into a positively charged amino acid in 14/52 cases (27%). Position 29 had fewer of such exchanges.

There were some subtype-specific patterns as well. Subtype D lost a negatively charged amino acid at position 25 in 100% of its sequences, while at position 29 the negatively charged amino acid was lost by 57% (4/7). Subtype C lost a negatively charged amino acid in only 11% (1/9) of its sequences, keeping D (8/9) at position 29. Subtypes B and C tended to lose negatively charged amino acids more at position 25 than at position 29.

The presence of positively charged amino acids at positions 11 and 25 are considered important markers for the X4 phenotype. We calculated the proportion of acquired positively charged amino acids at positions 11 and 25 versus other positions that were not subtype-specific or generally conserved ones (Table 1a). For all five subtypes, only 41/161 (25%) acquired positively charged amino acids were at positions 11/25, while 120/161 (75%) were at other positions (\( P = 0.0001 \), Wilcoxon signed rank test).

The influence of loss of negatively charged amino acids at the two negatively charged positions 25 and 29 on final charge was evaluated, including the extra effect of the negatively charged amino acids being substituted by positively charged amino acids. On average, the charge increase was +1.29 for all subtypes (Table 1b). However, new negatively charged amino acids were sometimes gained at positions other than 25 and 29, which affected the contribution of these positions to the total average charge increase for all subtypes, changing it from +1.29 to +0.95.

The 11/25 rule (positively charged amino acid at position 11 and neutral at position 25, or just positively charged amino acid at position 25) is often used to define the X4 phenotype. Here, 47% (26/55) of the X4 sequences followed this rule. We compared the impact of positions 11 and 25 on the overall gain in charge with the impact of other positions that were not subtype-specific or generally conserved ones (Table 1c). The impact of such positions on the overall charge gain was greater than the impact of positions 11 and 25 in each subtype, and statistically significant for all subtypes together (\( P = 0.0001 \), mixed repeated effect’s model).

In individual sequences other positions had an equal or greater impact on the overall charge increase compared with positions 11 and 25 in 100% of subtype A sequences, 63% of subtype B, 78% of subtype C, 86% of subtype D, 78% of CRF01_AE and 73% of the total sequences from all five subtypes, as can be deduced from Fig. 2.

In summary, we have shown that the increased charge of X4 versus R5 within each subtype was related to additions in most positions of the gp120 V3 region. In the R5 phenotype, we found both generally conserved and subtype-specific charged amino acid positions. The same ones were seen in X4, but with the addition of more charge and the loss of negative charge at positions that were conserved in R5. When positions 11 and 25 were compared with other positions regarding the impact on total charge of the sequences, we observed that other positions had a greater impact, decreasing the fundamental role of the 11/25 positions.

A random selection of one individually unique sequence from each individual provided the closest representation of the world HIV epidemic, resulting in a limited set of data per subtype. Therefore, we chose to include only the major subtypes. Still, we had only a few X4 sequences from some of the subtypes.

The source of the phenotyped sequences varied. In only a few cases, the sequences were obtained from the cells where
L. Pramanik and others

Fig. 2. Charged amino acid distributions in the X4 phenotype in different HIV-1 subtypes. All X4 sequences are displayed, together with an all consensus sequence from the HIV sequence database for comparison, listed according to subtype, followed by the GenBank accession number. Gaps in individual sequences are indicated with ‘~’ and gaps created by the alignment with ‘−’. Positions (11 and 25) are shaded, colours and charge are shown as in Fig. 1. Acquired positively charged amino acids are boxed in red, lost negatively charged amino acids are boxed in blue, and a combination of both are boxed in green. Column 11/25 shows what impact positions 11 and 25 have, while U (unspecified) what impact acquired and lost charged amino acids in positions other than 11 and 25 have on the total charge of each sequence. T stands for the total charge gained by 11/25 and U.
the phenotype was actually determined. All allocated R5X4 strains were excluded, because currently we cannot distinguish if these represent a mixture of R5 and X4 viruses with combined properties or virus strains that can use both co-receptors. In most other studies R5X4 are grouped together with X4 sequences.

Subtype C displayed a greater proportion of the R5 phenotype, as known before, and CRF01_AE of X4, which is new. The R5 sequences in all studied subtypes were very homogeneous, and the X4 sequences were heterogeneous, similar to findings in the M- and T-tropic strains (Chesebro et al., 1992), but the X4 sequences nevertheless preserved the positively charged general and subtype-specific positions that were observed in the R5 phenotype. Our data confirmed that positively charged amino acids are often found at positions 11 and 25 in the X4 phenotype (Resch et al., 2001), similar to findings in the SI phenotype (de Jong et al., 1992b; Fouchier et al., 1992). The 11/25 positions seemed to play a greater role in determining coreceptor use among subtype B viruses. Since subtype B viruses have been studied the most, this would naturally influence the overall interpretations regarding other subtypes. Our conclusion is that the total charge of V3 is important for X4 and that this is also the case for subtype B viruses.

We have characterized differences in charge at the amino acid level in the major HIV-1 subtypes. Our observations of less position-dependent acquisition of positive charge provide a simple and plausible explanation for how X4 evolves from R5. This advances our knowledge of HIV-1 phenotype–coreceptor interactions, which could improve computer-based models and may be of use in future drug and vaccine designs.

**Acknowledgements**

We gratefully acknowledge all the patients whose samples were used. Funds supporting this study were granted from the Karolinska Institute.

**Table 1. Amino acid source of charge increase in the X4 phenotype**

<table>
<thead>
<tr>
<th>Subtype</th>
<th>11/25</th>
<th>Other positions</th>
<th>Total</th>
<th>P-values*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>51</td>
<td>73</td>
<td>0.002</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>22</td>
<td>26</td>
<td>0.035</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>18</td>
<td>25</td>
<td>0.073</td>
</tr>
<tr>
<td>CRF01_AE</td>
<td>5</td>
<td>23</td>
<td>28</td>
<td>0.013</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>120</td>
<td>161</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. sequences</th>
<th>Total charge gain</th>
<th>Mean charge increase</th>
</tr>
</thead>
</table>

- **(a) Acquisition of positively charged amino acids.**
- **(b) Loss of negatively charged amino acids at positions 25 and 29.**
- **(c) Impact of the 11/25 rule versus other positions on total charge increase.**

*Wilcoxon signed rank test.*
†Unstructured mixed effect’s model for repeated measures. There was no significant influence by any particular subtype.
Institute to L. P. (an MD-PhD training grant) and A. E., and from SIDA to A. E.

References


