Original article

Identification and structural characterization of novel genetic elements in the HIV-1 V3 loop regulating coreceptor usage

Valentina Svicher[1], Claudia Alteri[1], Anna Artese[2], Jing Maria Zhang[3], Giosuè Costa[2], Fabio Mercurio[1], Roberta D'Arrigo[4], Stefano Alcaro[2], Giorgio Palù[5], Massimo Clementi[6], Maurizio Zazzi[7], Massimo Andreoni[8], Andrea Antinori[4], Adriano Lazzarin[9], Francesca Ceccherini-Silberstein[1], Carlo Federico Perno[1] on behalf of the OSCAR study group[†]*

[1]Department of Experimental Medicine and Biochemical Sciences, University of 'Tor Vergata', Rome, Italy
[2]Department of Pharmacobiological Sciences, University of Catanzaro 'Magna Graecia', Roccelletta di Borgia, Italy
[3]Department of Statistics, Yale University, New Haven, CT, USA
[4]INMI 'L Spallanzani', Rome, Italy
[5]Department of Histology, Microbiology and Medical Biotechnologies, University of Padua, Padua, Italy
[6]Laboratory of Microbiology and Virology, 'Vita-Salute San Raffaele' University, Milan, Italy
[7]Department of Molecular Biology, University of Siena, Siena, Italy
[8]Department of Public Health, University of 'Tor Vergata', Rome, Italy
[9]Department of Infectious Diseases, 'S Raffaele' Scientific Institute, Milan, Italy

*Corresponding author e-mail: cf.perno@uniroma2.it
[†]A list of members of the OSCAR study group can be found via Additional file 1

Background: The interaction between HIV-1 gp120 and CCR5 N terminus is critical for R5-virus entry and affects CCR5 antagonists' activity. Knowledge of how different genetic signatures of gp120 V3 domain effect the strength of this interaction is limited.

Methods: HIV-1 coreceptor usage was assessed in 251 patients using enhanced-sensitivity Trofile assay and V3 sequencing plus tropism prediction by Geno2pheno algorithm. Bayesian partitional model and recursive model selection have been used to define V3 genetic determinants correlated with different coreceptor usage. Gp120 interaction with CCR5 N terminus was evaluated by docking-analysis/molecular–dynamic simulations starting from the model described previously.

Results: Selected V3 genetic determinants (beyond known aminoacidic positions) significantly correlate with CCR5- or CXCR4-usage, and modulate gp120 affinity for CCR5 N terminus. This is the case for N5Y and N7K, absent in CCR5-using viruses and present in 4.5% and 6% of CXCR4-using viruses, respectively, and A19V, occurring in 2.6% of CCR5-using viruses and 22.0% of CXCR4-using viruses ($P=10^{-2}$ to 10^{-7}). Their presence determines a decreased affinity for CCR5 N terminus even stronger than that observed in the presence of the well-known mutation S11R (N5Y: -6.60 Kcal/mol; N7K: -5.40 Kcal/mol; A19V: -5.60 Kcal/mol; S11R: -6.70 Kcal/mol; WT: -6.90 Kcal/mol). N7K significantly increases the distance between V3 position 7 and sulphotyrosine at CCR5 position 14 (crucial for binding to gp120; from 4.22 Å to 8.30 Å), thus abrogating the interaction between these two important residues.

Conclusions: Key determinants for tropism within the V3 sequence, confirmed by structure- and by phenotypic-tropism, have been identified. This information can be used for a finer tuning of potential efficacy of CCR5-antagonists in clinical practice, and to provide molecular implications for design of new entry inhibitors.

Introduction

HIV-1 entry into host cells is a multistep process that requires coordinated interactions of the envelope glycoprotein gp120 with the CD4 receptor and with one of the chemokine receptors, CCR5 or CXCR4. Pure CCR5-tropic and pure CXCR4-tropic virus can use only the CCR5 and CXCR4 coreceptors to enter

target-cells, respectively, while dual-tropic virus can use both coreceptors [1].

The study of HIV-1 coreceptor usage is so far of central pathological importance because of its strict correlation with the rate of disease progression in HIV-1-infected individuals [2,3]. Determining HIV-1 coreceptor usage is also important since the CCR5 coreceptor has recently become the target of a new class of anti-HIV-1 drugs that specifically inhibit the entry of CCR5-tropic HIV-1 strains into the target cells by allosteric inhibition of the CCR5 coreceptor [4]. Maraviroc is the first approved CCR5 antagonist which entered clinical practice in 2007.

Among the different HIV-1 gp120 domains, the V3 loop is recognized as the primary determinant for HIV-1 coreceptor usage [5–7]. It is composed of three regions: a base, closely associated with the bridging sheet on the gp120, a flexible stem and a conserved-hairpin crown [6]. The V3 positions 11, 24 and 25 have been shown to be involved in binding with different coreceptors [8,9]. However, there is increasing evidence that these positions alone are not sufficient to fully understand mechanisms underlying different coreceptor usage. Despite several efforts, a systematic definition and characterization of V3 genetic determinants underlying different HIV-1 coreceptor usage is so far still missing.

In addition, recent studies demonstrated that the interaction of the V3 base with the N-terminal part of the CCR5 coreceptor is crucial for the entry of CCR5-using viruses into the target cells [7,10–12]. However, the effect of V3 mutations on the strength of this interaction has not been elucidated. Understanding this point is clinically relevant since an increased affinity of V3 for the CCR5 N terminus has been recently correlated with a reduced *in vitro* response to CCR5 antagonists, suggesting the ability of the CCR5-using virus to bind CCR5 despite the presence of the drug [13,14].

In this regard, using unique computational methods along with structural analysis and molecular dynamics simulations, this study is aimed at defining the V3 genetic determinants and the structural features underlying the ability of HIV-1 to use the CCR5 and/or CXCR4 coreceptors *in vivo*. This information can be adopted to get molecular data for a finer tuning of potential efficacy of CCR5-antagonists in clinical practice and for designing new candidates for entry inhibitors.

Methods

Study population

This study includes 323 plasma samples from 323 HIV-1 subtype-B-infected patients (all naive to maraviroc). For each patient, HIV-1 tropism was assessed both phenotypically, by the enhanced sensitivity version of Trofile, and genotypically, by V3 sequencing [15] and tropism prediction. The geno2pheno algorithm at a false-positive rate of 10% was used to infer HIV-1 coreceptor usage. Plasma samples for genotypic and phenotypic determination of HIV-1 tropism were collected at the same time-point. V3 sequences have been submitted to GenBank (accession number: HQ696137-HQ696387).

Statistical analyses

Mutation prevalence

Among the 323 samples characterized both genotypically and phenotypically for viral tropism, we selected 251 samples (192 reported as CCR5-tropic by both genotypic and phenotypic tests, and 59 reported as CXCR4-tropic or dual/mixed-tropic by the enhanced sensitivity version of Trofile and confirmed as CXCR4-tropic by genotypic tropism test). To assess the association of V3 mutations with different coreceptor usage, we calculated the prevalence of all the 311 amino acids (including the wild-type amino acids) found at the 35 V3 positions in 192 V3 sequences from CCR5-using viruses, and 59 V3 sequences from CXCR4-using viruses. We then performed χ^2 tests of independence to identify statistically significant differences in frequency between the two groups of patients. The significance was also confirmed by using a Monte Carlo procedure [16].

We used the Benjamini–Hochberg method at a false discovery rate of 0.05 [17] to identify results that were statistically significant in the presence of multiple hypothesis testing.

Mutation covariation

In the set of 251 V3 sequences, we exhaustively analysed patterns of pairwise interactions among V3 mutations associated with different coreceptor usage [18,19]. Specifically, for each pair of mutations and corresponding wild-type residues, Fisher's exact test was performed to assess whether co-occurrence of the mutated residues differed significantly from what would be expected under an independence assumption. Again, the Benjamini–Hochberg method was used to correct for multiple testing (false discovery rate of 0.05) [17].

Cluster analysis

In order to identify and summarize higher-order interactions of mutations, we transformed the pairwise phi correlation coefficients into dissimilarity values. Based on these pairwise dissimilarity values, a dendrogram was computed by hierarchical clustering. Finally, the stability of the resulting dendrogram was assessed from 100 bootstrap replicates [18,19].

Bayesian variable partition model and recursive model selection

Bayesian variable partition model and recursive model selection were carried out according to the methodology

described in Zhang *et al*. [20]. A detailed description of these models is reported in Additional file 2.

Structural analyses

Starting from the model of the CCR5 N terminus in complex with HIV-1 gp120-CD4 described by Huang *et al*. [7], we performed our structural analysis, generating all the analysed mutants by single-residue replacement.

Each complex was then placed in a cubic cell, with size adjusted to maintain a minimum distance of 10 Å to the cell boundary, and soaked with a pre-equilibrated box of water using the System Builder module of the Desmond package (Maestro–Desmond Interoperability Tools version 2.2; Schrödinger, New York, NY, USA) [21]. All overlapping solvent molecules were removed and an appropriate number of counter ions were added to maintain charge neutrality. In order to optimize the geometries, all the complexes were energy minimized, using OPLS2005 as force field [22,23]. Starting from the energy optimized geometry, all the analysed complexes were submitted to AutoDock Vina [24] docking simulations, using the CCR5 N terminus as ligand and the HIV-1 gp120-CD4 complex as receptor. For our simulations we adopted the following conditions: 100 allowed configurations per ligand; Gasteiger PEOE partial charges [25]; an exhaustiveness increased to 128; a cubic box of 70,196 $Å^3$, centered on CD2 atom of V3 I322 residue. For all the analysed complexes we selected the best pose, taking into account Huang's model geometric criteria [7]. In particular, we monitored the distance between N302 (corresponding to V3 position 7), R331, and I439 with the sulphotyrosine (Tys) at the CCR5 positions 10, 14 and with the Tyr at CCR5 position 15, respectively. Moreover we analysed the torsional angles of the CCR5 Tys residues and finally, for each analysed complex, we chose the configuration able to better reproduce these geometric parameters.

In addition, starting from the energy optimized geometry, all the analysed complexes were submitted to molecular dynamic simulations (MDSs) under the following conditions: recording interval equal to 10 ps; 5 ns of simulation time at 300 K; pressure set to 1 bar; OPLS2005 as force field [22,23] and SPC water molecules. All simulations were performed by Desmond package (Schrödinger) [21].

We performed the energy evaluation of electrostatic- and VdW-terms. In particular, for the coulombic term close electrostatic interactions were computed using an interpolation scheme with a cutoff radius of 9 Å and a long-range smooth particle mesh Ewald method with a tolerance of 1×10^{-9}. For VdW-term analysis we implemented the widely used Lennard–Jones non-bonded interactions.

After MDSs, for all the analysed complexes we carried out the interaction energy evaluation not considering the solvent contribution.

Results

Analytical pipeline for defining V3 genetic determinants underlying CCR5 and CXCR4 coreceptor usage

V3 genetic determinants related to different coreceptor usage were first inferred by using a subset of 251 V3 sequences from HIV-1 subtype-B-infected patients (all naive to the CCR5 antagonist maraviroc). For each patient, HIV-1 tropism was assessed both phenotypically by the enhanced sensitivity version of Trofile, and genotypically by V3-sequencing and tropism prediction using the geno2pheno algorithm set at a false-positive rate of 10%.

Such determinants were also assessed in an independent dataset composed of 315 V3 sequences with phenotypically determined tropism from 274 pure CCR5-using and 41 pure CXCR4-using HIV-1 isolates (all subtype B) collected from Los Alamos Database (Los Alamos, NM, USA). In this dataset, we applied both a Bayesian variable partition (BVP) model and a recursive model selection (RMS) procedure [20] to define V3 genetic determinants either individually or in network associated with CCR5- or CXCR4-usage, and to infer the detailed dependence structure among the interacting positions.

Finally, we illustrated the molecular basis of V3 determinants identified by using docking analysis, molecular dynamics simulations and free-energy decomposition analyses.

HIV-1 gp120 V3 determinants and their association with CCR5- or CXCR4-usage

Among the 311 V3 amino acids analysed, we identified 34 specific HIV-1 subtype-B genetic determinants at 19 V3 positions significantly correlated with CCR5 or CXCR4 coreceptor usage *in vivo* (Table 1).

Among the 29 mutations correlated with CXCR4-usage, 6 reside at the already known positions 11, 24 and 25 [8,9]. This is the case of S11K/R, E24K/R and E25K/R, completely absent or occurring in ≤1% of V3 sequences from CCR5-using viruses and with a prevalence ranging from 5.1% to 50.8% in V3 sequences from CXCR4-using viruses (*P*-values from 10^{-2} to 10^{-26}; Table 1).

Beyond these mutations, other 23 mutations at 15 V3 positions were significantly correlated with CXCR4-usage. The majority of them (16/23) were completely absent or present with a frequency of ≤1% in V3 sequences from CCR5-using viruses. Conversely, they had a significantly increased frequency in V3 sequences

Table 1. V3 amino acids associated with different coreceptor usage

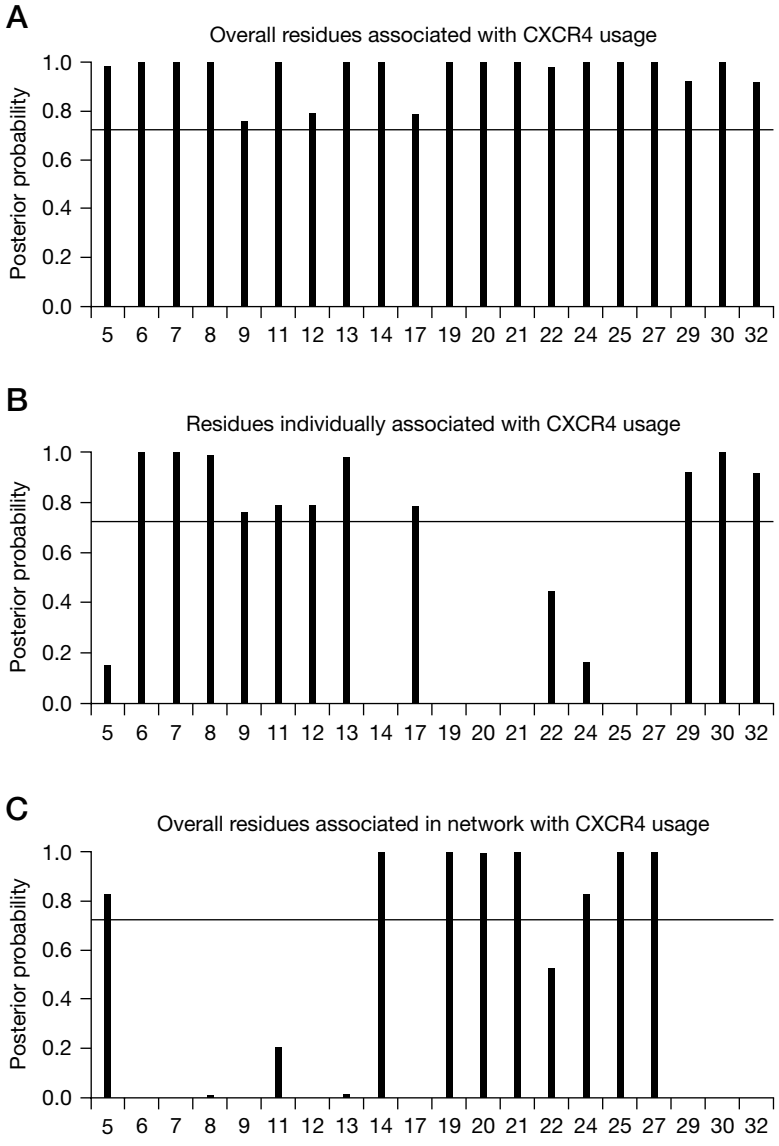
V3 amino acid[a]	Localization in V3[b] subdomain	Frequency[c] in R5-using viruses (*n*=192)	Frequency[c] in X4-using viruses (*n*=59)	*P*-value[d]
V3 determinants correlated with CXCR4 usage				
At the classical V3 positions 11, 24 and 25				
11K	Stem	0 (0.0)	6 (10.2)	**7.7×10^{-6}**
11R	Stem	0 (0.0)	30 (50.8)	**6.3×10^{-26}**
24K	Stem	1 (0.5)	3 (5.1)	1.0×10^{-2}
24R	Stem	1 (0.5)	5 (8.5)	**5.0×10^{-4}**
25K	Stem	2 (1.0)	10 (16.9)	**5.5×10^{-7}**
25R	Stem	0 (0.0)	9 (15.2)	**3.6×10^{-8}**
At other V3 positions				
5Y	Base	0 (0.0)	6 (10.2)	**7.7×10^{-6}**
N6GKTW	Base	2 (1.0)	5 (8.5)	**2.4×10^{-3}**
7K	Base	0 (0.0)	2 (3.4)	1.0×10^{-2}
7Y	Base	0 (0.0)	5 (8.5)	**4.6×10^{-5}**
7T	Base	0 (0.0)	3 (5.1)	**1.0×10^{-3}**
8A	Base	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
8I	Base	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
9K	Stem	2 (1.0)	6 (10.2)	**5.0×10^{-4}**
9I	Stem	1 (0.5)	4 (6.8)	**3.0×10^{-3}**
12L	Stem	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
12M	Stem	0 (0.0)	4 (6.8)	**3.0×10^{-3}**
13T	Crown	12 (6.2)	13 (22.1)	**4.0×10^{-3}**
13R	Crown	3 (1.6)	7 (11.9)	**4.0×10^{-3}**
14V	Crown	0 (0.0)	6 (10.2)	**7.7×10^{-6}**
17R	Crown	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
19S	Crown	3 (1.6)	8 (13.6)	**8.2×10^{-5}**
19V	Crown	5 (2.6)	13 (22.0)	**4.2×10^{-7}**
20Y	Crown	4 (2.1)	8 (13.6)	**3.0×10^{-3}**
21V	Stem	0 (0.0)	5 (8.5)	**4.6×10^{-5}**
27L	Base	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
30R	Base	0 (0.0)	3 (5.1)	**2.0×10^{-3}**
32K	Base	29 (15.1)	16 (27.1)	3.0×10^{-2}
32R	Base	5 (2.6)	6 (10.2)	1.0×10^{-2}
V3 determinants correlated with CCR5 usage				
At the classical 11, 24 and 25 positions				
11S	Stem	150 (78.1)	9 (15.2)	**1.9×10^{-18}**
25D	Stem	83 (43.2)	8 (13.6)	**3.4×10^{-5}**
At other V3 positions				
19A	Crown	176 (91.7)	28 (52.4)	**2.5×10^{-9}**
22A	Stem	132 (68.7)	29 (49.1)	6.0×10^{-3}
27I	Stem	23 (88.0)	22 (62.7)	**6.0×10^{-4}**

Frequencies represented as *n* (%). [a]All the 311 amino acids (including the wild-type amino acids) found at the 35 V3 positions have been analysed. The table reports only V3 amino acids significantly associated with CXCR4 or CCR5 usage. Underlined amino acids are the wild-type for HIV-1 B subtype gp120 consensus sequence; all other are mutations. [b]The V3 loop is composed of three regions: the base (V3 residues 3–8, 26–33) that is closely associated with the bridging sheet on the gp120 core, a flexible stem (V3 residues 9–12, 21–25), and a conserved hairpin crown (V3 residues 13–20; Huang *et al.*, [6]). [c]The analysis was performed on 192 samples reported as R5-tropic by both genotypic and phenotypic tests and on 59 samples reported as X4/DM-tropic by Trofile HS and X4-tropic by genotypic test. [d]Statistically significant differences were assessed by χ^2 tests of independence. *P*-values in bold are confirmed also after correction for multiple comparison by the Benjamini–Hochberg method using a false-positive rate of 0.05 (Benjamini & Hochberg, [17]).

from CXCR4-using strains (frequency ranging from 3.4% to 13.6%; *P*-values from 10^{-2} to 10^{-6}; Table 1).

We then studied the positions associated with CCR5-usage. A detailed analysis showed that only the wild-type amino acid at position 11 (serine [S]), 19 (alanine [A]) and 27 (isoleucine [I]) and the mutations E25D and T22A correlated with CCR5-usage (*P*-values from 10^{-4} to 10^{-18}; Table 1). Such limited number of determinants associated with CCR5 usage is fully consistent with the lower degree of V3 genetic

Figure 1. Posterior probabilities for each mutation to be associated with CXCR4 usage



(A) The graph reports the overall posterior probability for a mutation to be associated with CXCR4 usage, and the posterior probability to be **(B)** individually, and **(C)** associated in network with CXCR4 usage. Horizontal lines indicate posterior probability >0.75.

variability in CCR5-using viruses observed in this and other studies [6,26].

For a further definition of V3 determinants underlying different HIV-1 coreceptor usage, we applied in an independent dataset a unique procedure [20] to infer V3 positions that are either individually or in network associated with CCR5- or CXCR4-usage. Figure 1A shows the posterior probabilities >0.75 for each determinant to be associated with CXCR4 based on the BVP model. All the genetic markers identified in the above mentioned analysis were fully confirmed by this model. Among them, 14 V3-positions (5, 6, 7, 8, 11, 13, 14, 19, 20, 21, 24, 25, 27 and 30) showed a posterior probability to be associated with CXCR4-usage ≥0.99. These are strong predictors of CXCR4-usage. Mutations at the V3 positions 1, 3, 4, 10, 18, 26, 28, 31, 33, 34 and 35 had a posterior probability <0.15, thus indicating that they are not involved in mechanisms underlying different coreceptor usage.

Overall, the modelling on sequences drawn from Los Alamos Database is completely consistent with the analysis performed on V3 sequences from our

dataset. This supports the involvement of these genetic markers in mechanisms underlying coreceptor usage.

Dependence structure of interaction for HIV-1 coreceptor usage

Among the V3 positions reported in the previous paragraph, BVP model identified 11 genetic determinants (at V3 positions 6, 7, 8, 9, 11, 12, 13, 17, 29, 30 and 32) individually associated with CXCR4-usage (Figure 1B), and 9 V3 genetic determinants associated in network with CXCR4-usage (Figure 1C). These included the well-known V3 positions 24 and 25, and also the V3 positions 5, 14, 19, 20, 21, 22 and 27.

We thus applied the RMS procedure to infer the detailed dependence structure among the interacting positions, and we found that all these 9 positions associated in network with tropic status (5, 14, 19, 20, 21, 22, 24, 25 and 27), are tightly interacted with each other (with posterior probability >0.9; Figure 2A). No subgroup and substructures were found. Thus, this is a fully interacting group. This means that all these determinants cooperatively act to modulate HIV-1 coreceptor usage.

Specific associations among V3 determinants

Both BVP and RMS models highlighted the interaction among mutations at the V3 positions 5, 14, 19, 20, 21, 22, 24, 25 and 27. Thus, we investigated the complex interaction patterns of mutations at those positions.

Associations among V3 mutations

To identify significant patterns of pairwise correlations between mutations at the above mentioned positions, we calculated the binomial correlation coefficient (phi) and its statistical significance for each pair of mutations. A positive and statistically significant correlation between mutations at two specific positions ($0< \text{phi} <1$, $P<0.05$) indicates that these two positions co-evolve in order to confer an advantage to the virus; thus it indicates that the co-occurrence of mutations is not due to chance.

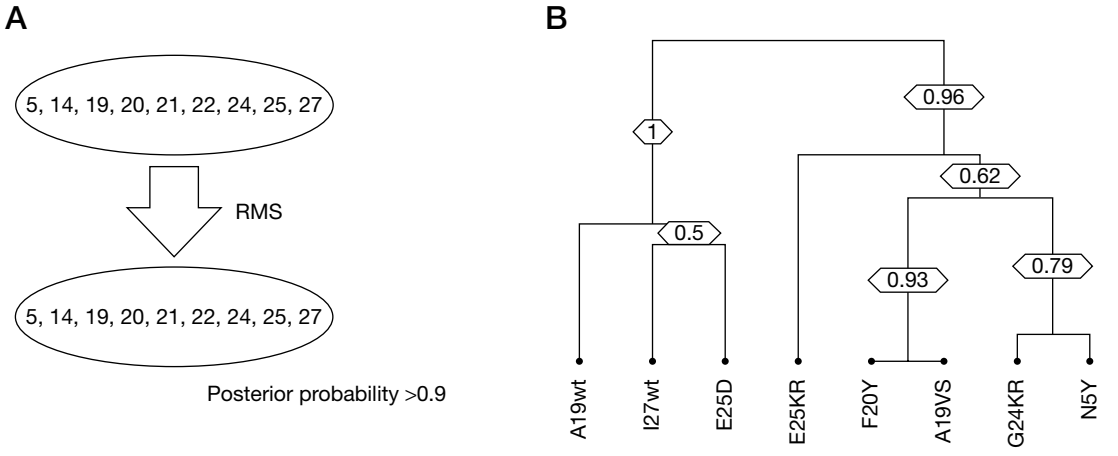
We found a strong and significant association of A19S/V with either F20Y (phi =0.44, *P*-value $=2.3\times10^{-5}$), or N5Y (phi =0.27, *P*-value $=1.2\times10^{-2}$). All these mutations are correlated with CXCR4-usage. A strong negative association was observed between A19V/S and E25D (this last correlated with CCR5 usage; phi =-0.22, *P*-value $=3.8\times10^{-3}$). In particular, the copresence of E25D and A19S/V was observed only in 2 out of 251 (0.7%) V3 sequences analysed, suggesting an antagonism between these mutations.

Conversely, E25D was specifically associated with the wild-type amino acids at the V3 positions 19 ($A19A_{wt}$; phi =0.17, *P*-value $=2.8\times10^{-2}$) and 27 ($I27I_{wt}$; phi =0.22, *P*-value $=3.6\times10^{-3}$). They might represent a favourable genetic background for the acquisition of E25D.

Clusters of correlated mutations

Because pairwise analysis suggested that most V3 genetic determinants are associated with specific evolutionary pathways underlying different coreceptor

Figure 2. V3 mutational patterns characterizing HIV-1 tropism



(A) Detection of a detailed mutation interaction structure for V3 position associated in network with CXCR4 usage. **(B)** The dendrogram, obtained from average linkage hierarchical agglomerative clustering, shows significant clusters involving V3 mutations. Length of branches reflects distances between mutations in the original distance matrix. Bootstrap values, indicating the significance of clusters, are reported in the boxes. The analysis was performed on 251 V3 sequences. RMS, recursive model selection procedure.

usage, we performed average linkage hierarchical agglomerative cluster analysis to investigate this hypothesis in more detail.

The topology of the dendrogram (Figure 2B) highlights the existence of two distinct clusters of V3 determinants preferentially associated with either CCR5- or CXCR4-usage. In particular, a larger cluster of V3 determinants associated with CXCR4-usage was observed. This cluster involved the V3 determinants correlated with CXCR4-usage: N5Y, A19V/S, F20Y and G24K/R (bootstrap value =0.62; Figure 2B). This whole cluster was strongly linked to E25K/R (bootstrap value =0.96; Figure 2B).

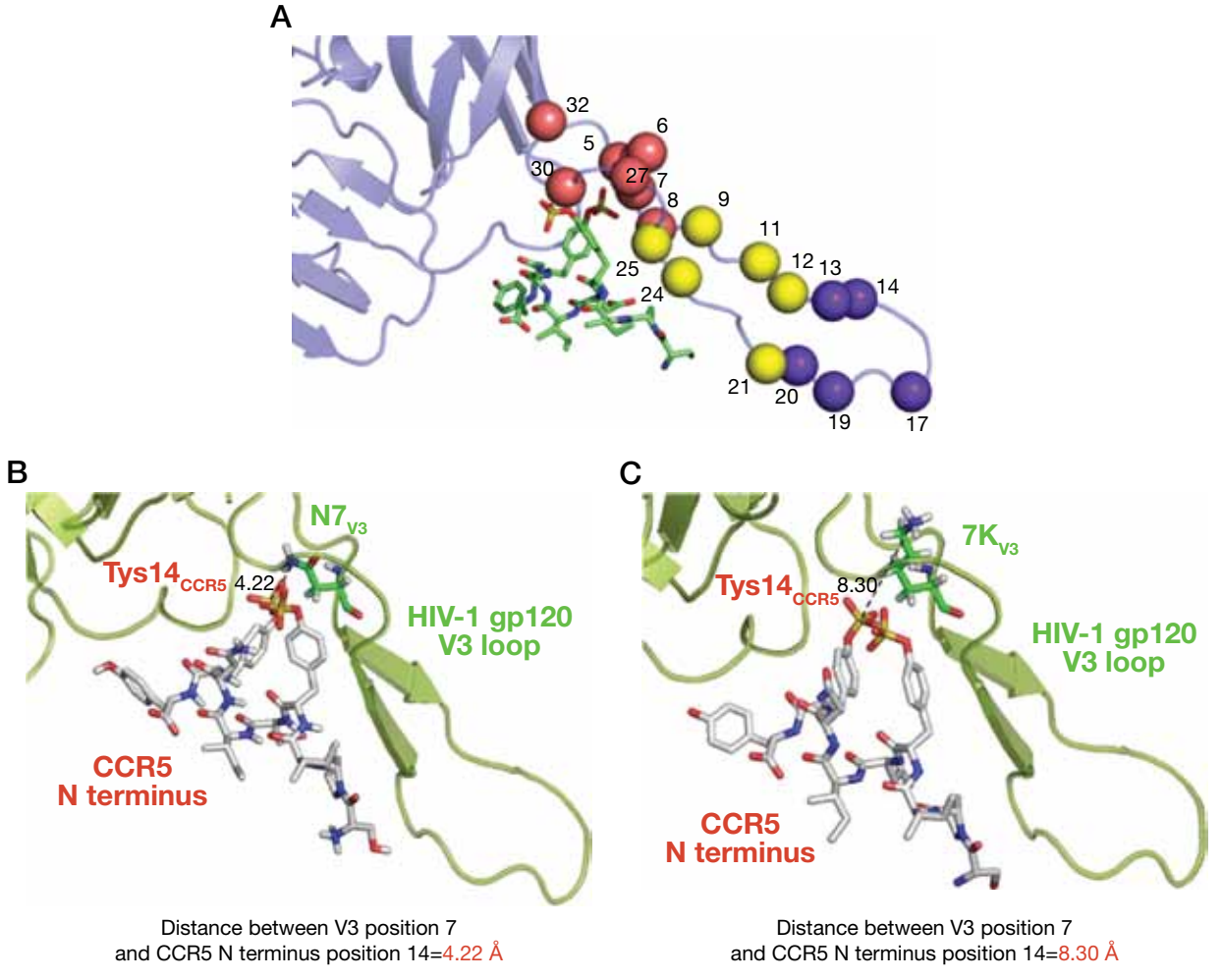
Conversely, a strong cluster was formed by E25D, and the wild-type amino acids A and I at V3 positions 19 and 27 (bootstrap value =1.00), respectively (Figure 2B). The involvement of different wild-type amino acids in this cluster again confirms the higher degree of conservation of CCR5-using isolates.

Structural characterization of V3 genetic determinants correlated with different coreceptor usage

Localization of V3 residues in the three-dimensional structure of HIV-1 gp120 V3 loop

V3 genetic determinants involved in coreceptor usage are mainly localized in the base (N5Y, N7K/Y/T, T8A/I, I27L, I30R, Q32K/R) or in the stem (R9I/K, S11K/R/S, I12L/M, Y21V, T22A, G24R, E25D/K/R) of the V3 loop (Figure 3A). Of note, V3 residues

Figure 3. Localization of V3 residues involved in coreceptor usage



(A) The localization of V3 residues involved in coreceptor usage in the structure of HIV-1 gp120 V3 domain is shown. The V3-loop complexed with CCR5 N terminus is shown. The HIV-1 gp120 is represented as light blue cartoon. Residues in the V3 base, stem and crown are shown as salmon, yellow and blue spheres, respectively. The CCR5 N terminus is represented as green carbon sticks. The panels **(B)** and **(C)** show the best pose after docking simulation of wild-type and N7K mutant, respectively, reporting the distance, expressed in Å, between V3 position 7 and CCR5 N terminus position 14. HIV-1 gp120 is represented as light green cartoon; N terminus of CCR5 is shown as green carbon sticks. Molecular docking was performed starting from the model described in Huang *et al.* [7].

individually associated with different coreceptor usage are localized on either side of the V3 base, and in the N-terminal part of the V3 stem encompassing positions 9 and 12. This supports the role of the entire V3 base and the N-terminal part of the V3 stem in regulating coreceptor binding.

Molecular basis of V3 mutations revealed by docking analysis and free energy calculations

The V3 base is known to interact with the N terminus of the CCR5 coreceptor, and this interaction has been shown to be critical for the entry of CCR5-using viruses into the target cells [7,10–12]. Thus, starting from the model of the CCR5 N terminus in complex with HIV-1 gp120-CD4 described in Huang *et al*. [7], we conducted docking analysis and molecular dynamics simulations to investigate the effect of V3 determinants (identified in previous analyses) on the strength of this interaction.

Beyond the classical position 11 and 25, we focused our attention also on the determinants localized in the base of the V3 loop. We also included in our analysis A19V, localized in the stem of V3-loop, since covariation analysis revealed a strong negative interaction among mutations at position 19 and 25. In our wild-type model, all the structural parameters reported in Huang *et al*. [7] were fully reproduced by docking simulations.

The free energy analyses of wild-type and mutated complexes showed that both the well-known mutations S11R and E25R are related to a decreased binding affinity for CCR5 N terminus compared to the HIV-1 subtype B wild-type gp120 (Figure 4), thus supporting that this structural model may accurately capture the current knowledge regarding the role of position 11 and 25 in mechanisms underlying coreceptor usage. Beyond S11R and E25R, also the mutations N5Y, N7K, A19V and Q32K, all correlated with CXCR4-usage, determined a decreased binding affinity for CCR5 N terminus (Figure 4). In particular, N5Y, A19V and N7K determined a less favourable energy profile, even lower than that observed in presence of the well known mutation S11R (Figure 4). This further confirms the involvement of such determinants in CXCR4-usage. No change in binding affinity was observed in presence of T8A.

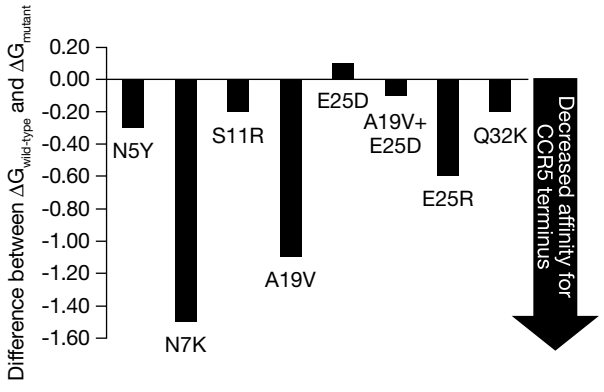
Consistently with its association with CCR5-usage, the mutation E25D was found linked with an increased binding affinity for the CCR5 N terminus compared to the wild-type gp120 (Figure 4). Thus, in presence of E25D the interaction between gp120 and the CCR5 N terminus is stronger. This increase in binding affinity is abolished when E25D is present along with A19V. This result confirms the negative association that we observed between these 2 mutations.

The free energy decomposition analyses, carried out after our molecular dynamics simulations for the wild-type and the mutants, highlighted that the V3 mutations primarily affect the electrostatic interactions between CCR5 and the HIV-1 subtype B gp120. In particular, we observed most unfavourable energy profiles related to N7K, S11R, A19V and Q32K mutations if compared to the wild-type sequence (Additional file 3). Conversely, the increased binding affinity for CCR5 N terminus observed in presence of E25D is mainly related to the van der Waals interactions between gp120 and CCR5 (Additional file 3).

There is evidence that the gp120 residues N302 (corresponding to V3 position 7), R331 and I439 directly interact with the sulphotyrosine (Tys) at the CCR5 positions 10 and 14, and with the Tyr at CCR5 position 15, respectively [7]. In particular, the interaction between N7 and Tys14 in CCR5 are critical for a proper binding between gp120 and the CCR5 N terminus [7]. Thus, in order to better unravel the molecular mechanisms underlying different coreceptor usage, we also monitored the distance between these gp120 and CCR5 residues. We found that N7K increased the distance between the V3 position 7 and the Tys at CCR5 N terminus position 14 (Figure 3B and 3C). This long distance abrogates the interaction between these two important residues. Again, this is fully consistent with the strongest decrease in binding affinity observed in presence of N7K and with its independent association with CXCR4-usage. We also observed that Q32K increased the distance between the gp120 residue 439 and the Tys at CCR5 position 15.

Conversely and consistent with its ability to increase the binding affinity for the CCR5 N terminus, E25D

Figure 4. Impact of V3 mutations on binding affinity for CCR5 N terminus



The graph reports the variation of the free energy differences (ΔΔG) in kcal/mol between the wild-type ($\Delta G_{wild\text{-}type}$) and the mutated ($\Delta G_{mutant}$) complexes of HIV-1 subtype B gp120. Docking analysis was performed starting from the model of gp120-CD4 complexed with CCR5 N terminus described by Huang *et al.* [7]. Results were confirmed by molecular dynamic simulations.

reduced the distance between the V3 position 7 and 31 with the Tys at the CCR5 N terminus position 10 and 14, respectively, thus increasing the chance of proper binding between CCR5 and its ligand gp120 at V3 loop.

Overall, genotypic and structural data consistently confirm the presence of specific pathways of V3 genetic determinants underlying different coreceptor usage.

Discussion

This study provides a systematic approach for a better definition of V3 genetic determinants underlying different coreceptor usage *in vivo*. It shows that other V3 genetic determinants, beyond the classical positions 11, 24 and 25, can modulate HIV-1 subtype B usage of CCR5 or CXCR4 coreceptor.

The genetic determinants, that we identified, were confirmed in two independent datasets by combining different and unique mathematical models. The molecular basis underlying the mechanisms of action of such determinants were further unveiled by docking analysis, molecular dynamics simulations and free-energy calculations. Previous studies have not addressed this point before.

We identified and characterized a number of V3 determinants significantly associated with CXCR4-usage. Mutations N5Y, N7K/Y/T, T8A/I, S11K/R, I14V, Y21V, I27L and I30V are absent in V3-sequences from CCR5-using viruses, and had a posterior probability to be associated with CXCR4-usage >0.99, thus indicating that they are strongly predictive of CXCR4-usage.

The correlation of N5Y, N7K/Y/T and T8A/I with CXCR4-usage is consistent with a previous study showing that deletions of V3 residues 5 to 7 within the dual-tropic strain HIV-189.6P can selectively ablate CCR5-tropism and increase HIV-1 susceptibility to the CXCR4 antagonist AMD3100 [12,27]. Another study has shown that mutations at positions 7 and 8 determine a 50% reduction in CCR5-dependent cell-to-cell fusion [28].

These mutations are all localized at the base of the V3 loop, known to directly interact with the CCR5 N terminus [7]. Our free-energy evaluation showed that the presence of either N5Y or N7K decreases the binding affinity of the CCR5 N terminus for the V3 loop. In addition, docking analysis showed that N7K increases the distance between V3 position 7 and sulphotyrosine at CCR5-position 14 (crucial for binding to gp120), thus abrogating the interaction between these two important residues. Thus, structural analysis confirms the role of these two mutations in impairing HIV-1 ability to use CCR5 coreceptor, and supports that V3 base contributes to stabilize a structure indispensable for CCR5-utilization.

In our structural analysis, T8A did not affect gp120 binding affinity for the CCR5 N terminus. It is conceivable that T8A can mediate the interaction with the CXR4 coreceptor *via* the loss of the N-linked glycosylation site at the V3 position 6-8.

Moreover the free- energy decomposition highlighted the crucial role of the electrostatic term for the CCR5 N terminus affinity *versus* the V3 loop. In particular, we found that all the mutations resulting in the acquisition of a positive charge, such as N7K, S11R and Q32K, are related to a remarkably unfavourable profile for CCR5 binding. The importance of positive charge at position 11 is well known [8,9], while this is the first study highlighting the role of positive charges at position 7 and 32 in mechanisms underlying CXCR4-usage. Similarly, also A19V, localized in V3 stem, strongly decreases V3 binding affinity for the CCR5 N terminus. This supports the involvement of different V3 domains in modulating the interaction with the CCR5 N terminus.

By using a unique procedure based on Bayesian statistical modelling, we also delineated the complex interactions among these mutations. In particular, RMS showed that the V3 positions 5, 14, 19, 20, 21, 22, 24, 25 and 27 tightly interact with each other, thus forming a full interaction group significantly and strongly associated with CXCR4 usage. The presence of position 25 in this cluster is consistent with a recent study showing that basic amino acid substitutions at position 25 (R or K) alone are not sufficient to confer efficient CXCR4 usage [29]. Again, these results support that positions 11 and 25 alone are not sufficient to fully understand the mechanisms underlying different coreceptor usage, and highlight the association between multiple amino acid changes in V3 and efficient CXCR4 usage.

Our structural analysis showed that E25D is associated with an increased binding affinity for the CCR5 N terminus. Recent studies have shown that an increased affinity of V3 for the CCR5 N terminus correlates with a reduced *in vitro* response to CCR5 antagonists [13,14], thus highlighting a correlation between the genetic and structural determinants underlying CCR5 usage and response to CCR5 antagonists *in vivo*. In particular, specific mutations in the V3 loop have been shown to confer *in vitro* resistance to vicriviroc [14].

Overall, further studies are necessary to investigate the effect of such determinants on: mechanisms underlying coreceptor usage in HIV-1 non-B subtypes and recombinants, virological response to maraviroc in clinical cohorts, and on disease progression. Regarding this last point, we found that (among all V3 determinants associated with CXCR4 usage identified) N5Y and T8A/I were associated with lower CD4 count, (103 ±115 cells/μl with N5Y versus 349 ±188 cells/μl without N5Y, P=0.03, and 37 ±24 cells/μl with T8A/I versus 302 ±175 cells/μl without T8A/I, P=0.02), suggesting their potential ability to interfere with HIV-1 cytopathicity [30], and thus further supporting the importance of conducting such studies on larger cohorts.

In conclusion, new genetic determinants of tropism within the V3 domain have been detected by phenotype–genotype correlation and confirmed by structural analysis. This information can be used for a finer tuning of potential efficacy of CCR5 antagonists in clinical practice, and to get molecular implications for design of new entry inhibitors.

Acknowledgements

We gratefully thank Jun S Liu for suggestions in statistical analysis. This work was financially supported by grants from CHAIN, Collaborative HIV and Anti-HIV Drug-Resistance Network, Integrated Project number 223131, funded by the European Commission Framework-7 Program, and from the Italian Ministry of Health (CUP:E81J10000000001).

These data have been presented, in part, at the *XVII Conference on Retrovirus and Opportunistic Infections (CROI)*, 15–19 February 2010, San Francisco, CA, USA*; 8th European HIV Drug Resistance Workshop*, *From Basic Science to Clinical Implications*, 16–19 March 2010, Sorrento, Italy*; International HIV and Hepatitis Virus Drug Resistance Workshop*, 10–12 June 2010, Dubronvik, Crotia.

Disclosure statement

The authors declare no competing interests.

Additional files

Additional file 1: A complete list of centres and members participating in the OSCAR programme can be found at http://www.intmedpress.com/uploads/documents/AVT-11-OA-1990_Svicher_Add_file1.pdf

Additional file 2: An explanation of the Bayesian variable partition model and recursive model selection can be found at http://www.intmedpress.com/uploads/documents/AVT-11-OA-1990_Svicher_Add_file_2.pdf

Additional file 3: A table displaying energy evaluation values, after molecular dynamics simulations, can be found at http://www.intmedpress.com/uploads/documents/AVT-11-OA-1990_Svicher_Add_file3.pdf

References

1. Berger EA, Doms RW, Fenyö EM, *et al*. A new classification for HIV-1. *Nature* 1998; **391:**240.
2. Waters L, Mandalia S, Randell P, *et al*. The impact of HIV tropism on decreases in CD4 cell count, clinical progression, and subsequent response to a first antiretroviral therapy regimen. *Clin Infect Dis* 2008; **46:**1617–1623.
3. Regoes RR, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol* 2005; **13:**269–277.
4. MacArthur RD, Novak RM. Reviews of anti-infective agents: maraviroc: the first of a new class of antiretroviral agents. *Clin Infect Dis* 2008; **47:**236–241.
5. Hoffman TL, Doms RW. HIV-1 envelope determinants for cell tropism and chemokine receptor use. *Mol Membr Biol* 1999; **16:**57–65.
6. Huang CC, Tang M, Zhang MY, *et al*. Structure of a V3-containing HIV-1 gp120 core. *Science* 2005; **310:**1025–1028.
7. Huang CC, Lam SN, Acharya P, *et al*. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* 2007; **317:**1930–1934.
8. Fouchier RA, Groenink M, Kootstra NA, *et al*. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* 1992; **66:**3183–3187.
9. De Jong JJ, De Ronde A, Keulen W, *et al*. 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* 1992; **66:**6777–6780.
10. Cormier EG, Tran DN, Yukhayeva L, *et al*. Mapping the determinants of the CCR5 amino-terminal sulfopeptide interaction with soluble human immunodeficiency virus type 1 gp120-CD4 complexes. *J Virol* 2001; **75:**5541–5549.
11. Farzan M, Chung S, Li W, *et al*. Tyrosine-sulfated peptides functionally reconstitute a CCR5 variant lacking a critical amino-terminal region. *J Biol Chem* 2002; **277:**40397–40402.
12. Nolan KM, Jordan AP, Hoxie JA. Effects of partial deletions within the human immunodeficiency virus type 1 V3 loop on coreceptor tropism and sensitivity to entry inhibitors. *J Virol* 2008; **82:**664–673.
13. Nolan KM, Del Prete GQ, Jordan AP, *et al*. Characterization of a human immunodeficiency virus type 1 V3 deletion mutation that confers resistance to CCR5 inhibitors and the ability to use aplaviroc-bound receptor. *J Virol* 2009; **83:**3798–3809.
14. Ogert RA, Ba L, Hou Y, *et al*. Structure–function analysis of human immunodeficiency virus type 1 gp120 amino acid mutations associated with resistance to the CCR5 coreceptor antagonist vicriviroc. *J Virol* 2009; **83:**12151–12163.
15. Svicher V, D'Arrigo R, Alteri C, *et al*. Performance of genotypic tropism testing in clinical practice using the enhanced sensitivity version of Trofile as reference assay: results from the OSCAR Study Group. *New Microbiol* 2010; **33:**195–206.
16. Hope ACA. A simplified Monte Carlo significance test procedure. *J R Stat Soc [Ser A]* 1968; **30:**582–598.
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JR Statist Soc B* 1995; **57:**289–300.
18. Svicher V, Sing T, Santoro MM, *et al*. Involvement of novel human immunodeficiency virus type 1 reverse transcriptase mutations in the regulation of resistance to nucleoside inhibitors. *J Virol* 2006; **80:**7186–7198.
19. Svicher V, Aquaro S, D'Arrigo R, *et al*. Specific enfuvirtide-associated mutational pathways in HIV-1 Gp41 are significantly correlated with an increase in CD4(+) cell count, despite virological failure. *J Infect Dis* 2008; **197:**1408–1418.
20. Zhang J, Hou T, Wang W, Liu JS. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proc Natl Acad Sci U S A* 2010; **107:**1321–1326.
21. Bowers KJ, Chow E, Xu H, *et al*. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC '06. Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. New York: ACM 2006.
22. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 1996; **118:**11225–11236.

23. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 2001; **105:**6474–6487.

24. Trott O, Olson JA. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; **31:**455–461.

25. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* 1980; **36:**3219–3228.

26. Bozek K, Thielen A, Sierra S, *et al*. V3 loop sequence space analysis suggests different evolutionary patterns of CCR5- and CXCR4-tropic HIV. *PLoS ONE* 2009; **4:**e7387.

27. Yang X, Tomov V, Kurteva S, *et al*. Characterization of the outer domain of the gp120 glycoprotein from human immunodeficiency virus type 1. *J Virol* 2004; **78:**12975–12986.

28. Suphaphiphat P, Thitithanyanont A, Paca-Uccaralertkun S, Essex M, Lee TH. 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* 2003; **77:**3832–3837.

29. Huang W, Frantzell A, Toma J, *et al*. Defining evolutionary pathways and genetic barriers to CXCR4-mediated entry by HIV type 1. *Antivir Ther* 2010; **15 Suppl 2:**A019.

30. Svicher V, Alteri C, Artese A, *et al*. Specific genetic signatures in V3 base can modulate co-receptor usage *in vivo*, the interaction with neutralizing antibodies, and HIV-1 cytopathic effect. *1st International HIV and Viral Hepatitis Virus Workshop*. 8–12 June 2010, Dubrovnik, Croatia. Abstract 12.

Accepted 15 February 2011; published online 27 July 2011