BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 4 2013, pages 451–460
doi:10.1093/bioinformatics/btt002

Structural bioinformatics

Advance Access publication January 6, 2013

Detecting and understanding genetic and structural features in HIV-1 B subtype V3 underlying HIV-1 co-receptor usage

Mengjie Chen[1,†], Valentina Svicher[2,†], Anna Artese[3], Giosuè Costa[3], Claudia Alteri[2], Francesco Ortuso[3], Lucia Parrotta[3], Yang Liu[4], Chang Liu[1], Carlo Federico Perno[2,*], Stefano Alcaro[3] and Jing Zhang[1,4,*]

[1]Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA, [2]Department of Experimental Medicine and Surgery, University of Tor Vergata, Rome, [3]Department of Health Sciences, University ''Magna Græia'' of Catanzaro, Campus ''S. Venuta'', Viale Europa, 88100, Catanzaro (CZ), Italy and [4]Department of Statistics, Yale University, New Haven, CT 06511, USA

Associate Editor: Martin Bishop

ABSTRACT

Motivation: To define V3 genetic elements and structural features underlying different HIV-1 co-receptor usage *in vivo*.

Results: By probabilistically modeling mutations in the viruses isolated from HIV-1 B subtype patients, we present a unique statistical procedure that would first identify V3 determinants associated with the usage of different co-receptors cooperatively or independently, and then delineate the complicated interactions among mutations functioning cooperatively. We built a model based on dual usage of CXCR4 and CCR5 co-receptors. The molecular basis of our statistical predictions is further confirmed by phenotypic and molecular modeling analyses. Our results provide new insights on molecular basis of different HIV-1 co-receptor usage. This is critical to optimize the use of genotypic tropism testing in clinical practice and to obtain molecular-implication for design of vaccine and new entry-inhibitors.

Contact: jing.zhang.jz349@yale.edu or cf.perno@uniroma2.it

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on September 29, 2012; revised on December 18, 2012; accepted on December 29, 2012

1 INTRODUCTION

HIV is the etiological agent of the acquired immunodeficiency syndrome, affecting ~40 million people worldwide (World Health Organization, http://www.who.int/hiv/en/). HIV is an enveloped virus with a single-stranded RNA genome that mainly infects CD4+ lymphocytes and macrophages. HIV entry is a complex and intricate process that facilitates the delivery of HIV-1 genome into the host cell. The only viral surface protein, envelope (Env), is composed of a trimer of gp120 and gp41 heterodimers. To enter the target cell, HIV requires a coordinated interaction of the envelope glycoprotein gp120 with the CD4 receptor and with one of the chemokine receptors, mainly CC chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4). HIV-1 strains can be phenotypically classified according to virus ability to use the CCR5 and/or CXCR4 co-receptor. Thus, pure R5-tropic and pure X4-tropic virus can use only the CCR5 and CXCR4 co-receptors to enter the target cell, respectively (Berger *et al.*, 1998), whereas dual-tropic virus can use both co-receptors. HIV-1 co-receptor usage is so far of central pathological importance. It has been shown that R5-tropic viruses are generally responsible for the establishment of the initial infection and predominate in the majority of newly HIV-1-infected patients, whereas the use of the CXCR4 co-receptor is generally observed in more advanced stages of disease, and it has been associated with a more rapid CD4 decline and progression to AIDS (Berger *et al.*, 1998; Regoes and Bonhoeffer, 2005). The determination of HIV-1 co-receptor usage is also critical, as the CCR5 co-receptor 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 co-receptor (Dorr *et al.*, 2005; Princen and Schols, 2005; Regoes and Bonhoeffer, 2005). Maraviroc is the first approved CCR5 antagonist that entered clinical practice in 2007. Its administration is so far recommended only in patients with pure R5-tropic viruses.

Among the different domains of HIV-1 gp120, the V3 loop is so far the focus of intense research efforts because it is the primary determinant for HIV-1 co-receptor usage (Hoffman and Doms, 1999; Huang *et al.*, 2005, 2007a). Nevertheless, a comprehensive map of V3 genetic markers and structural features underlying, individually or in network, the pure or dual usage of CCR5 and CXCR4 co-receptor is still missing. In particular, limited information is available regarding V3 genetic determinants underlying dual tropism. This is critical, as different studies have shown that high proportion of patients harbor dual-mixed tropic viruses, with prevalence ranging 12–15% in drug-naïve and 20–50% in treatment-experienced patients (Church *et al.*, 2008; Huang *et al.*, 2007b; Lihana *et al.*, 2009; Moreno *et al.*, 2009; Shepherd *et al.*, 2008).

In this light, using unique computational methods along with structural analysis and docking 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 CXCR4 co-receptors.

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

2 METHODS AND MATERIALS

2.1 The dataset and data processing

We downloaded HIV-1 type B V3 sequences with phenotypic determination of viral tropism (mainly based on recombinant-virus entry assay on U87-CD4+-CCR5+-/CXCR4+-expressing cells) from http://www.hiv.lanl.gov/ (Kuiken *et al.*, 2010) using Sequence Search Interface (http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) where we specified the download option 'co-receptor' to be 'only CCR5', 'only CXCR4' or 'only R5X4'. After downloading the data, we performed a phylogenetic analysis. After phylogenetic analysis, only one sequence per patient was conserved. This resulted in a dataset consisting of 390 V3 sequence: 277 were associated with a CCR5 co-receptor usage, 41 with CXCR4 co-receptor usage and 72 with dual tropism. All the sequences data are provided in Supplementary Material. All DNA sequences were aligned using Basic Local Alignment Search Tool and then translated into amino acid sequence based on three-letter codons.

2.2 Bayesian partition on dual usage of co-receptor model

We developed a Bayesian Partition on Dual Usage of Co-receptor Model (BPDUCM) to define V3 genetic determinants either independently or interactively associated with the usage CCR5 co-receptor only, CXCR4 co-receptor only or dual of CCR5/CXCR4 co-receptor.

In analysis based on three datasets—CCR5 only, CXCR4 only and dual usage—suppose there are N_t sequences from CXCR4-using viruses, Nu sequences from CCR5-using viruses and N_w sequences from dual-using viruses. Each sequence is of p-residues long (for V3 here, $p=35$). Let $X=\{X_1,\ldots, X_p\}$ be the observations of sequences. X_j is a column vector that contains N ($N=N_t+N_u+N_w$) observations at the jth position. Residue type X_{ij} at position j of the ith sequence can be one of L_j-possible amino acids. We also define dataset indicator $Y=\{Y_1,\ldots, Y_N\}$, which represents the status of co-receptor usage of each sequence, that is, $Y_i=0$ if the ith sequence is from a CCR5-using viruses, $Y_i=1$ if it is from a CXCR4-using viruses and $Y_i=2$ if it is from a dual-using viruses.

The N_t sequences are assumed to be independent and identically distributed (IID) observations of the variables $X=\{X_1,\ldots, X_p\}$ from population using only CXCR4 co-receptor; the Nu sequences are IID observations of these variables from population using only CCR5 co-receptor; and the N_w sequences are IID observations of these variables from population using dual of CXCR4/CCR5 co-receptor.

Our goal is to describe the complicated relationship between the sequence observations (X) and the dataset indicator (Y). Basically, we partitioned the p ($p=35$) positions into K groups ($K=10$ in Figure 1a, G1 ~ G10 represent Group 1 to Group 10) according to their relationship to Y. Each of the K groups represent one relationship between X and Y. Then we calculate the posterior probabilities of different partitions (partition p positions into K groups). We denote $I=(I_1, I_2,\ldots, I_p)$ as the group indicator. $I_j=k$ ($j=1,\ldots,$ p and $k=1,\ldots,$ K) means the jth position is partitioned into the kth group. Given Y, we observe X and want to infer I when p and K are fixed. The likelihood is P (X|I, Y), and the posterior probability is

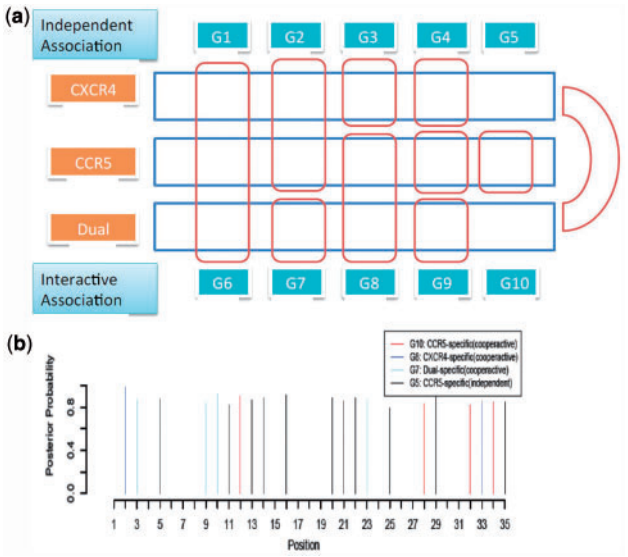

Fig. 1. (**a**) BPDUCM. (**b**) The posterior probabilities for each mutation in domain V3 to be associated with co-receptor usage inferred from BPDUCM. The graph reports the overall posterior probabilities for a mutation to be cooperatively associated with CCR5 usage (red, G10 in a) and CXCR4 usage (dark blue, G8 in a), dual usage of CXCR4/CCR5 usage (light blue, G7 in a) and independently associated with CCR5 usage (gray, G5 in a)

P (I|X, Y), which is proportional to the previous P (I|Y) times the likelihood P (X|I, Y). We also assume that I is independent of Y previously, therefore, P (I|Y) = P (I). A Markov chain Monte Carlo algorithm was then designed to sample from the posterior distribution of I so as to infer which positions are associated with the co-receptor usage status, with the following proposals: randomly choose one position and change its group membership I_i.

Regarding the possible independently or interactively associations among three datasets, we proposed a BPDUCM: G1 ~ G5 for variables (positions) in X independently associated with Y and G6 ~ G10 for variables interactively associated with Y. Details of BPDUCM model can be found in Supplementary Material.

2.3 Recursive model selection

After BPDUCM found interacting positions, we applied recursive model selection (RMS) (proposed in our previous article, Zhang *et al.*, 2010, and now has been widely applied in a number of HIV and HBV studies, Svicher *et al.*, 2011a,b,c,d; Svicher *et al.*, 2012; Zhang *et al.*, 2012) to infer the detail structure of interactions. In the aforementioned BPDUCM, variables in G6 ~ G10 are fitted into a 'fully saturated' model. However, in practice, it is desirable to know the detailed conditional independence structure within each interaction group. RMS is to first infer among two classes of cruder models, that is, the chain-dependence model and the V-dependence model, and then recursively apply this strategy until the data does not support more detailed models.

2.4 GRID-based pharmacophore model application

The GRID-based pharmacophore model (GBPM) method has been described in a recent publication (Ortuso *et al.*, 2006). In the

GRID (Goodford, 1985) calculations, the lone pairs, the tautomeric hydrogen atoms and torsion angles, relative to the sp3 oxygen atoms and the amide atoms, have been allowed to be settled on the basis of the probe influence, whereas the coordinates of all the other atoms have been considered rigid (directive MOVE = 0). Default values have been used for the other parameters.

2.5 Docking simulations

Starting from the model of the CCR5 N-terminus in complex with HIV-1 gp120-CD4 described in the study by Huang *et al.* (2005), we performed our CCR5 analysis generating all the analyzed mutants by single-residue replacement.

To test the accuracy and reliability of our docking approach, we firstly performed CCR5 re-docking calculation on the wild-type (WT) HIV-1 gp120 CD4 sequence. We used AutoDock Vina docking software (Trott and Olson, 2010) defining the CCR5 N-terminus as ligand and the HIV-1 gp120-CD4 complex as receptor. For our simulations, we adopted the following conditions: (i) 100 allowed configurations per ligand; (ii) Gasteiger partial equalization of orbital electronegativities partial charges (Gasteiger and Marsili, 1980); (iii) an exhaustiveness increased to 128; and (iv) a cubic box of 70196Å3, centered on CD2 atom of V3 I25 residue.

After single-residue replacement, each mutated 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 (Bowers *et al.*, 2006; Desmond version 2.2, Schrödinger, Inc., Portland, OR, USA). All overlapping solvent molecules were removed, and an appropriate number of counter ions were added to maintain charge neutrality. To optimize the geometries, all the complexes were energy minimized, using OPLS2005 as force-field (Jorgensen *et al.*, 1996; Kaminski *et al.*, 2001).

After docking procedure validation, starting from the energy optimized geometry, all the analyzed complexes were submitted to AutoDock Vina (Trott and Olson, 2010) docking simulations using the same aforementioned protocol.

For all the CCR5 studied complexes, we selected the best pose taking into account the Huang *et al.*'s (2007a) model geometric criteria. In particular, we monitored the distance between N7, R31 and I397 with the sulfotyrosine (Tys) at the CCR5 positions 10, 14, and with the Tyr at CCR5 position 15, respectively. Moreover, we analyzed the torsional angles of the CCR5 Tys residues, and finally, for each analyzed complex, we chose the configuration able to better reproduce these geometric parameters.

By contrast, for all the CXCR4 analyzed complexes, we selected the best energy scored pose, as no experimental structural data were available.

Finally, all the selected complexes best docking poses were submitted to optimization and thermodynamic evaluation using MacroModel version 7.2 (MacroModel version 7.2, Schrödinger Inc., Portland, OR, USA, 1998–2001; Mohamadi *et al.*, 1990) in the following conditions: AMBER* as force field (McDonald and Still, 1992), 10 000 iterations of minimization and the GB/SA water implicit model (Still *et al.*, 1990) to take into account the solvation effect. All the obtained conformations were evaluated in terms of hydrogen bonds (HBs) and VdW contacts by means of Maestro graphical interface (Maestro Graphics User Interface, version 9.8, Schrödinger, LLC). The component interaction analysis was performed starting from the optimized best pose of all studied complexes in the following conditions: (i) AMBER* as force field; (ii) GB/SA water implicit solvation model; (iii) dielectric constant equal to 1; and (iv) a binding pocket defined considering gp120 residues within 12 Å from the co-receptor (Maestro Graphics User Interface, version 9.8, Schrödinger, LLC).

All 3D figures were performed using PyMOL graphics and modeling package version 1.3 (The PyMol Molecular Graphics System, version 1.3, Schrödinger, LLC—DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org).

Our molecular modeling calculations were performed on a Linux cluster with fifteen dual-Xeon processor nodes.

3 RESULTS

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

With V3 sequences downloaded from Los Alamos Database (http://www.hiv.lanl.gov/; Kuiken *et al.*, 2010), there is a compendium of 390 DNA sequences for HIV-1 B subtype patients V3 domain, among which 277 use CCR5 co-receptor only, 41 use CXCR4 co-receptor only and 72 use dual of CCR5/CXCR4 co-receptor (see Section 2). We proposed and applied BPDUCM (see Section 2, Fig. 1a) on the aforementioned 390 HIV-1 B subtype gp120 V3 amino acid sequences. Figure 1b shows the posterior probability from this model for their association with usage of different co-receptors. We identified and characterized a number of V3 determinants significantly associated with the usage of those three cases (posterior probability > 0.8). Among them, V3 genetic determinants 5, 11, 13, 14, 16, 20, 21, 22, 25 and 29 independently associated with CCR5 co-receptor. Notably, the results suggested some cooperative associations: V3 genetic determinants 12, 28, 32 and 34 interacted with each other modulating HIV-1 ability to use the CCR5 specifically (corresponding to G10 in Fig. 1, localization of these four V3 residues is reported in Supplementary Figures S1 and S2). Then we applied RMS (see Section 2) to infer the dependent structure of interactions. RMS showed that these four positions could be further divided into two independent subgroups: 12 and 28, and 32 and 34. Tables 1 and 2 show the detail distributions of WT and mutant types in these two groups. CCR5 clearly has different distribution compared with CXCR4 and dual. The WT amino acids at those positions are associated with CCR5 usage, whereas the majority of mutations at those positions are associated with CXCR4-usage.

For position 12 and 28, from Table 1, no mutation at 28 was observed in CCR5, whereas G28E was observed in both CXCR4 (2.4%) and dual (2.8%). However, in CXCR4, only single mutation G28E was observed, whereas double mutation I12F + G28E shows up in dual. This indicates an interesting epistasis effect between these two mutations that we investigated in more detail by our computational analysis described in next section. For positions 32 and 34, single mutations Q32K (Q32R) and H34Y and the double mutation Q32K + H34Y are observed

Table 1. Frequencies of mutation combinations at 12 and 28

Mutation combination	CXCR4 (41)	Dual (72)	CCR5 (277)	*P*-value
I12 + G28	82.93%	83.30%	95.31%	0.000397
I12L + G28	4.88%	1.40%	0.00%	0.026082
I12M + G28	0.00%	0.00%	0.40%	NA
I12V + G28	7.32%	12.50%	4%	0.021742
I12 + G28E	2.44%	0	0	0.014024
I12F + G28E	0	2.80%	0	0.011802
I12R + G28	2.40%	0	0	0.014024
I12T + G28	0	0	0.40%	NA

NA, not applicable.

Table 2. Frequencies of mutation combinations at 32 and 34

Mutation combination	CXCR4 (41)	Dual (72)	CCR5 (277)	*P*-value
Q32 + H34	53.66%	56.94%	78%	5.52E-05
Q32 + H34Y	7.32%	8.33%	7.22%	0.949329
Q32K + H34	12.20%	26.39%	7.22%	7.21E-06
Q32R + H34	14.63%	4.17%	0.72%	2.50E-06
Q32K + H34Y	9.76%	2.78%	4.69%	2.45E-01
Q32R + H34Y	2.44%	0.00%	0.36%	NA
Q32 + H34F	0	0	0.36%	NA
Q32 + H34S	0	0	0.36%	NA
Q32H + H34	0	1.39%	0	0.109263
Q32K + H34F	0	0.00%	0.36	NA
Q32K + H34N	0	0	0.36%	NA
Q32L + H34Y	0	0	0.36%	NA

in all of the three groups of patients, indicating their ability to adopt CCR5, CXCR4 and dual.

On the basis of the statistical results, we further performed molecular modeling and structure analysis to shed light on the molecular and structural aspects of their epistasis impact on co-receptor–specific usage. The thermodynamic details are reported in Table 3, and the obtained results are described in the following. We have also analyzed the interactions, in terms of HBs and VdW contacts, of both co-receptors in all the studied complexes (Supplementary Tables S1–S36) to evaluate the contribution of single gp120 residues in the complex stabilization.

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

Starting from the model of the CCR5 N-terminus in complex with HIV-1 gp120-CD4 described in Huang *et al.* (2005), we performed docking and thermodynamics evaluation analyses to investigate the impact of V3 determinants on the strength of CCR5–gp120 interaction. In our WT model, all the structural parameters reported in Huang *et al.* (2007a) were fully reproduced by docking simulations (Supplementary Fig. S3).

By contrast, for CXCR4 complexes, we adopted the 2K05 (Veldkamp *et al.*, 2008) PDB (www.rcsb.org) model of the co-receptor complexed with the chemokine stromal cell-derived factor 1. In particular, we used Huang's model as receptor, and we applied the same protocol adopted for CCR5. Specifically in our experiments, we used the gp120 core obtained from a CCR5-dependent HIV-B subtype isolate YU2. Consistent with this, in gp120 WT sequence, CCR5 was associated to a more favorable free energy of complexation if compared with CXCR4 (−67.2 kcal/mol CCR5 versus −40.1 kcal/mol CXCR4).

Table 3. Evaluation of the free energy of complexation calculated for all gp120 optimized complexes in the presence of both co-receptors CCR5 and CXCR4

Model	ΔG_{bind} (kcal/mol)	
	CCR5	CXCR4
WT	−67.2	−40.1
S11R	−76.3	-2.17×10^4
I12L	−60.7	−134.0
I12M	−55.4	−147.0
I12R	−52.0	−75.6
I12T	−55.7	−209.0
I12V	−69.1	−85.8
E25K	−60.5	−122.0
E25R	−60.9	−80.8
G28E	−51.2	−111.0
I12F + G28E	−57.0	−57.9
Q32H	−59.3	−45.2
Q32K	−61.1	−99.5
Q32R	−63.1	−68.9
H34N + Q32K	−57.8	−218.0
H34Y	−58.0	−199.0
H34Y + Q32K	−71.9	−65.4
H34Y + Q32R	−57.8	−164.0

GBPM application in co-receptors recognition step: With the aim to select the most involved residues in the co-receptors recognition step, our GBPM computational approach (Alcaro *et al.*, 2010a) was applied onto the WT gp120 CCR5 and CXCR4 optimized best poses. The GBPM method is an example of unbiased pharmacophore model approaches generation based on the GRID approach (Goodford, 1985). Pharmacophore models are useful tools for drug discovery and lead optimization processes, as they allow collection of most relevant structural features of biological active molecules. One of the first studies to demonstrate the impact of computational methods on drug design, published by von Itzstein *et al.* (1993), highlighted the energetically favorable site for an amino or guanidine group in the active site of the influenza virus neuraminidase using the GRID program. This approach guided the design and modification of a transition-state analogue lead compound, ultimately resulting in the drug Relenza.

In our analysis first of all, we used the well-known GRID software (Goodford, 1985) to highlight the co-receptors crucial moieties. In particular O- and SO probe molecular interaction fields (MIFs), representing the sp2 anionic oxygen and the pyramidal sulfur, respectively, well recognized CCR5 and CXCR4 sulfotyrosines at position 14 and 107, respectively (Supplementary Figs S4 and S5). At CCR5–gp120 interface, area residues

K175, R3, G24, E25, I26, D29, I30, R31, P395, P396, I397 and R398 were identified using three different probes: a generic hydrophobic (DRY), an HB acceptor (O) and an HB donor (N1). Because the obtained global energy minimum GRID points (Emin) were ranked in a wide range of values, graphical analysis of the GRID maps was carried out by considering, for each probe, an energy threshold (Ecut) equal to 60% of the gp120–CCR5 complex Emin, as reported in our previous work (Alcaro *et al.*, 2010b).

The aforementioned GBPM protocol was adopted for WT gp120 CXCR4 optimized best pose using the same probes and indicating as mainly involved in the co-receptor recognition the residues reported in Table 4.

Specifically, Supplementary Figures S6 and S7 and Table 4 show the gp120 amino acids selected by GBPM using DRY, N3+ (cationic sp3 nitrogen), O and N1 probes at energy minimum. In particular, at CCR5 interface, E25, I30, R31 and P396 were well identified, whereas in CXCR4 recognition residues, C173, S177, R3 and A394 were resulted crucial. Specifically, DRY MIFs highlighted the hydrophobic residues I30 and A394 in CCR5 and CXCR4 complexes, respectively, whereas by analyzing the O probe MIFs R31, C173 and R3 were found to be pivotal for CCR5 and CXCR4 interactions. In gp120–CCR5 optimized best pose E25 side-chain carboxylate and P396 backbone oxygens resulted well defined by N3+ and N1 probes. At gp120–CXCR4 interface area, the HB donor probe was able to identify S177 side-chain hydroxyl, justifying the ability of these residues to produce HBs.

Such an interesting qualitative analysis prompted us to quantitatively estimate the role of the key identified amino acids in the co-receptors recognition step. In particular, a component interaction analysis was performed to estimate the average value of the single four residues contributions for CCR5 and CXCR4 interaction energy. All the obtained data are reported in Tables 5 and 6.

Single mutations at 12 and 28: As shown in Table 3, it is evident that the major contribution of V3 positions 12 and 28 to co-receptor usage is mainly because of the WT amino acids. Indeed, single mutations at these positions were found almost exclusively in CXCR4-using or dual-tropic viruses (see Table 1). This is the case of mutations at V3 position 12. Single mutations found at V3 position 12 were related to CCR5 unfavorable affinity profile if compared with the WT sequence (−67.2 kcal/mol WT, −60.7 kcal/mol I12L and −52.0 kcal/mol I12R), whereas they remarkably increased CXCR4 affinity (−40.1 kcal/mol WT, −134.0 kcal/mol I12L and −75.6 kcal/mol I12R). The only mutation at position 12 with high prevalence in dual-tropic viruses resulted I12V, even if it was found slightly predominant in CXCR4-using viruses (7.3% in CXCR4-using strains, 12.4% in dual-tropic strains and 4% in CCR5-using strains). Thus, consistent with this, I12V showed an increased CXCR4 N-terminus binding affinity for gp120 with respect to CCR5 (−69.1 kcal/mol for CCR5 and −85.8 kcal/mol for CXCR4, respectively).

These energetic profiles resulted well correlated with the contacts analysis for both the co-receptors. In particular, in CCR5 mutated complexes, in presence of I12L and I12R substitutions, the co-receptor lost pivotal HB with R31 and P396 with respect to the WT sequence; moreover, in I12R complex, the number of good contacts with the crucial glutamate at position 25 was observed remarkably reduced if compared with the WT (Supplementary Tables S3 and S5).

Interestingly, in I12V optimized configuration, even if CCR5 resulted able to establish an higher number of HB with respect to the WT, it lost just those with R31 and P396 and showed a decreased number of contacts with E25, completely abrogating the interaction with R31 (Supplementary Table S7). By contrast for CXCR4, in all mutated complexes at position 12, the co-receptor resulted able to establish additional contacts with the crucial residues identified by GBPM analysis if compared with the WT. Specifically, in I12L and I12V configuration, CXCR4 was well stabilized by some HB with C173, R9 and A394, whereas in I12R complex, we noticed an increased number of HB with respect to the WT, and in particular, with the arginine at position 9 (Supplementary Tables S21, S23 and S25). Our geometric observations were found in perfect agreement with the component interaction analysis either for CCR5 or for CXCR4, where the role of the selected residues at the co-receptors interface area was further confirmed.

Table 4. GBPM analysis performed using the WT models of CCR5 and CXCR4 gp120 optimized complexes

Residue	CCR5	CXCR4
C89		*
C173		*[a]
P174		*
K175	*	*
S177		*[a]
R3	*	
P4		*
N6		*
N7		*
R9		*[a]
G24	*	*
E25	*[a]	*
I26	*	*
I27		*
G28		*
D29		*
I30	*[a]	*
R31	*[a]	*
Q32		*
E348		*
P375		*
R377		*
K379		*
Q380		*
A394		*[a]
P395	*	*
P396	*[a]	*
I397	*	
R398	*	*
Q400		*

The asterisks point out gp120 residues highlighted by the GBPM approach at the co-receptor interface areas using DRY, N3+, O and N1 probes (E_{cut} equal to 60% of E_{min}).
[a]Indicates the amino acids identified by GBPM at energy minimum.

Table 5. CCR5 component interaction analysis carried out for all gp120 optimized complexes

Model	E25	I30	R31	P396	Average
WT	−32.46	−21.29	−25.35	−37.70	−29.18
S11R	−23.05	−25.00	−35.93	−36.46	−30.11
I12L	−17.73	−23.93	−42.39	−1.81	−21.59
I12M	−13.54	−8.20	−23.82	−26.26	−17.96
I12R	−20.35	−15.66	−14.27	−40.44	−22.68
I12T	−40.60	−16.53	−10.24	−2.81	−17.55
I12V	−37.42	−5.47	−34.30	−44.67	−30.47
E25K	−10.49	−30.10	−9.57	−41.64	−22.95
E25R	−1.41	−7.85	−12.67	−3.58	−6.38
G28E	−44.01	−0.91	−28.39	−0.99	−18.58
I12F + G28E	−3.37	−38.34	−18.58	−40.97	−25.32
Q32H	−19.00	−8.05	−9.11	−25.09	−15.31
Q32K	−17.53	−5.56	−1.62	−1.52	−6.56
Q32R	−36.41	−27.22	−16.93	−3.63	−21.05
H34N + Q32K	−29.46	−18.57	−22.86	−31.43	−25.58
H34Y	−5.35	−20.96	−3.02	−31.16	−15.37
H34Y + Q32K	−17.98	−44.23	−31.84	−30.37	−31.05
H34Y + Q32R	−8.04	−25.61	−24.08	−39.06	−24.20

The indicated residues were highlighted by GBPM approach at energy minimum. The interaction energy values are expressed in kcal/mol.

Table 6. CXCR4 component interaction analysis carried out for all gp120 optimized complexes

Model	C173	S177	R9	A394	Average
WT	−9.32	−14.77	−34.78	−25.39	−21.07
S11R	−30.87	−22.46	−26.42	−45.96	−31.43
I12L	−39.40	/	−33.79	−44.87	−29.51
I12M	−30.86	−37.56	−12.25	−40.69	−30.34
I12R	−10.78	−9.07	−45.28	−42.69	−26.96
I12T	−31.89	−36.02	−14.26	−29.80	−27.99
I12V	−11.20	−27.39	−33.85	−37.89	−27.58
E25K	/	/	−46.59	/	−11.65
E25R	−40.73	−41.71	−38.99	/	−30.36
G28E	−40.73	−13.04	−33.45	−7.56	−23.70
I12F + G28E	−6.31	−31.08	−11.99	−38.38	−21.94
Q32H	−40.93	−9.86	−25.59	−28.74	−26.28
Q32K	−32.19	−27.21	−30.53	−25.82	−28.94
Q32R	−25.05	−12.30	−27.54	−36.13	−25.25
H34N + Q32K	−45.48	−40.26	−8.08	−34.04	−31.97
H34Y	−16.33	−27.29	−33.00	−40.95	−29.39
H34Y + Q32K	−0.95	−33.90	−17.84	−43.12	−23.95
H34Y + Q32R	−43.77	−30.38	−34.30	−3.66	−28.03

The indicated residues were highlighted by GBPM approach at energy minimum. The interaction energy values are expressed in kcal/mol.

At position 28, single-mutation G28E was exclusively found in CXCR4 (Table 1), which is consistent with the thermodynamic data, as it decreased CCR5 affinity if compared with the WT sequence (−67.2 kcal/mol WT and −51.2 kcal/mol G28E), whereas it remarkably increased CXCR4 affinity (−40.1 kcal/mol WT and −111.0 kcal/mol G28E). Such an energetic profile was consistent with the analysis of the interactions. In fact, in CCR5 mutated configuration, the co-receptor showed an increased global number of contacts, but it missed the pivotal HB with P396 and was able to establish only one VdW contact with E25, thus resulting less stabilized (Supplementary Table S10). Also the component interaction analysis highlighted the different contribution of P396, amino acid of the gp120 C4 domain, in WT and G28E complexes (−37.30 kcal/mol WT versus −0.99 kcal/mol G28E), confirming the geometric result. By contrast, in CXCR4 component interaction analysis (Table 6), the most important single contribution resulted from residue C173, which was able to establish several VdW contacts with the co-receptor (Supplementary Table S28). Moreover, CXCR4 showed two crucial HB with S177, not present in the WT complex, confirming its increased affinity versus gp120 with respect to the WT.

3.3 Epistasis effect of I12F and G28E and tradeoff in dual usage of CCR5 and CXCR4

Concerning the double mutation at 12 and 28, I12F and G28E show interesting epistatic effect. From Table 1, single-mutation G28E was observed only in CXCR4 (but not in CCR5 or dual), whereas double-mutation I12F + G28E was observed only in dual. As shown in our thermodynamic analysis, the single-mutation G28E was found to improve CXCR4 affinity, remarkably decreasing CCR5 affinity versus gp120 (thus, it cannot be observed in either CCR5 or dual, as dual requires ability to bind with both CXCR4 and CCR5). Such an observation is fully consistent with the exclusive presence of G28E alone on CXCR4-using viruses. However, as reported in Table 3, in presence of the double-mutation I12F + G28E, both co-receptors energetic profiles resulted favored if compared with the single-mutation G28E ones (for CCR5, −57.0 kcal/mol I12F + G28E, −51.2 kcal/mol G28E; for CXCR4, −57.9 kcal/mol I12F + G28E, −111.0 kcal/mol G28E).

Analyzing CCR5 contacts in presence of the double mutation, the pivotal HB with I30, R31 and P396 resulted abrogated, whereas only one VdW interaction with E25 was found (Supplementary Table S11). By contrast, the thermodynamic improvement of CXCR4 N-terminus could be rationalized by two additional HB with R9, residue not involved in the recognition of the co-receptor in the WT configuration (Supplementary Table S29).

More specifically, this effect can be observed also in the co-receptors component interaction analysis. Indeed, the V3 residue 28 is directly involved in the interface area between V3 and CXCR4 N-terminus (Table 4). As indicated in CXCR4 component interaction analysis, the presence of G28E was strongly associated to the interaction with C173 in gp120 C2 domain, whereas this effect was abrogated by the co-presence of I12F with G28E (Table 6). On the other hand, CCR5 component interaction analysis indicated that when I12L is added to G28E, the average contribution of the four identified residues was similar to that of the WT complex, but specifically, C173 and R9 resulted energetically unfavorable if compared with the WT (Table 5).

Therefore, the double-mutation I12F + G28E observed in dual is a tradeoff between CCR5 and CXCR4 in the sense that its binding affinity with CCR5 is better than single-mutation G28E but worse than WT, whereas its binding affinity with CXCR4 is

better than WT but worse than G28E. Thus, its ability to bind with both CCR5 and CXCR4 turns out to be better than both WT and single-mutation G28E. This ability to bind with both co-receptors is achieved by the epistatic effect between I12F and G28E.

Single mutations at 32 and 34: A different situation was found for V3 positions 32 and 34. Although in this case, the contribution to pure CCR5 co-receptor usage is also driven by the WT amino acids and all the single mutations (with a prevalence >1%, Table 2) found at V3 positions 32 and 34 determined a strong increase in CXCR4 affinity (−40.1 kcal/mol WT, −68.9 kcal/mol Q32R, −99.5 kcal/mol Q32K and −199.0 kcal/mol H34Y) except Q32H (−45.2 kcal/mol); the decreases in CCR5 affinity did not result as large as in I12R and G28E cases if compared with WT (−67.2 kcal/mol WT, −63.1 kcal/mol Q32R, −61.1 kcal/mol Q32K, −59.3 kcal/mol Q32H, −58.0 kcal/mol H34Y, −52.0 kcal/mol I12R and −51.2 kcal/mol G28E).

The reduced CCR5 N-terminus stabilization in presence of the single-mutations Q32H, Q32K, Q32R and H34Y was further validated by our analysis of contacts. In particular, we observed that the co-receptor lost crucial HB with I30, R31 and P396 residues (Supplementary Tables S12, S13, S14 and S16), with a remarkable lack of VdW contacts with E25, especially in Q32H and H34Y mutated complexes. Moreover, the single contribution of the aforementioned amino acids resulted significantly decreased if compared with the WT sequence, and in particular, I30 and R31 for Q32H (−21.29 kcal/mol I30WT versus −8.05 kcal/mol I30Q32H; −25.35 kcal/mol R31WT versus −9.11 kcal/mol R31Q32H), I30, R31 and P396 for Q32K (−21.29 kcal/mol I30WT versus −5.56 kcal/mol I30Q32K; −25.35 kcal/mol R31WT versus −1.62 kcal/mol R31Q32K; and −37.70 kcal/mol P396WT versus −1.52 kcal/mol P396Q32K), P396 for Q32R (−37.70 kcal/mol P396WT versus −3.63 kcal/mol P396Q32R) and E25 and R31 for H34Y (−32.46 kcal/mol E25WT versus −5.35 kcal/mol E25H34Y and −25.35 kcal/mol R31WT versus −3.02 kcal/mol R31H34Y).

By contrast, CXCR4 N-terminus recognition of the glycoprotein in presence of Q32H, Q32R and H34Y single mutations indicated more favorable configurations of the co-receptor if compared with the WT complex, in particular when the hystidine at position 34 was replaced by a tyrosine. Specifically, such a thermodynamic improvement could be because of the additional HB established by CXCR4 with S177 in presence of Q32H/K/R substitutions, as reported in Supplementary Tables S30–S32. Indeed in the most favored H34Y mutated complex, the co-receptor was found to enhance its interaction network, making one HB with S177 and resulting well stabilized by two pivotal HB with R9 (Supplementary Table S34).

The component interaction analysis further supported our geometric considerations, showing an improved average contribution in all mutated complexes if compared with the WT, with particular evidence in presence of the single-mutation H34Y (Table 6).

3.4 Different epistasis effects for Q32K + H34Y and Q32R + H34Y

Although each of the single mutations Q32K and H34Y mildly decrease CCR5 affinity (−67.2 kcal/mol WT, −61.1 kcal/mol Q32K and −58.0 kcal/mol H34Y) and strongly increase CXCR4 affinity (−40.1 kcal/mol WT, −99.5 kcal/mol Q32K and −199.0 kcal/mol H34Y), the double-mutation Q32K + H34Y was found to increase the affinity in both CCR5 and CXCR4 (Table 3).

In CCR5 recognition, such a double mutation was related to an increased number of good contacts if compared with the WT, even if the co-receptor lost HB with I30, R31 and P396 (Supplementary Table S17). However, we observed that, in presence of both Q32K and H34Y substitutions, CCR5 N-terminus was well stabilized by additional HB with two arginine residues at position 3 and 398, respectively, emphasizing the crucial role of the electrostatic term in the co-receptor recognition step. Moreover, despite the reduced HB network of CCR5 with I30 and R31, their single contribution resulted much more improved with respect to the WT sequence (Table 5).

Also CXCR4 thermodynamic improvement, obtained in presence of Q32K + H34Y double mutation, was rationalized by our geometric observations. Specifically, with respect to the WT configuration, we revealed an additional HB with S177 and several VdW interactions with A394. Our component interaction analysis validated such a finding, highlighting the strong contribution of both residues in CXCR4 affinity versus the glycoprotein (−33.90 kcal/mol S177 and −43.12 kcal/mol A394).

A different behavior was observed for the Q32R + H34Y double mutation (Table 3). Indeed, although the single substitution Q32R (similar as Q32K and H34Y) was associated to an unfavorable CCR5 gp120 recognition (−67.2 kcal/mol WT, −63.1 kcal/mol Q32R and −58.0 kcal/mol H34Y) and improved CXCR4 affinity (−40.1 kcal/mol WT, −68.9 kcal/mol Q32R and −199.0 kcal/mol H34Y), the addition of H34Y to Q32R caused a further decrease in CCR5 affinity if compared with the single-mutation Q32R (−63.1 kcal/mol Q32R and −57.8 kcal/mol Q32R + H34Y) and a strong improvement in CXCR4 affinity for gp120 (−68.9 kcal/mol Q32R and −164.0 kcal/mol Q32R + H34Y). Thus, the addition of H34Y to Q32R reinforced the effect on CCR5 and CXCR4 affinity observed in presence of the single Q32R.

In particular, the analysis of CCR5 N-terminus contacts in the best optimized configuration in presence of the double mutation revealed that, even if the number of contacts was increased with respect to the Q32R single mutation, the co-receptor lost its HB with I30 and almost abrogated its interactions with E25 (Supplementary Table S18). Such an observation was validated by our component interaction analysis, where the single contribution of both residues, and in particular of the glutamate, resulted unfavorable with respect to the WT complex.

Conversely, CXCR4 N-terminus enhanced stabilization in presence of Q32R + H34Y double mutation seemed related to the single contribution of S177, rather than to the global interactions network (Supplementary Table S36). In fact, by comparing the geometric analyses when Q32R mutation was present alone and associated with H34Y, we observed that CXCR4 N-terminus established one HB with S177 in both the complexes. However, the co-receptor best energetic profile in Q32R + H34Y optimized configuration was related to a stronger contribution of the serine at position 177 if compared with the Q32R single substitution (−12.30 kcal/mol S177Q32R versus −30.38 kcal/mol S177Q32R + H34Y).

4 DISCUSSION

In this article, we described a systematic approach for a better definition of V3 genetic determinants underlying different co-receptor usage *in vivo*. This approach enabled us to capture sophisticate associations of V3 genetic determinants among the usage of co-receptors CCR5 and CXCR4 and even dual usage of them. This information is critical to optimize the use of genotypic tropism testing in clinical practice and to obtain molecular implication for design of vaccine and new entry inhibitors. In particular, the current algorithms for tropism prediction classify the virus with respect to its co-receptor usage genetic using statistical learning methods based on specific V3 sequence feature and, in particular, on amino acid substitutions observed at specific V3 positions. The genetic determinants identified and characterized in this study can be used to improve predictive performances of such algorithms.

We proposed an advanced model to detect mutation associated with different co-receptor usage and to infer detailed interaction structures among these mutations.

We focused on three cases: viruses using CCR5 co-receptor, CXCR4 co-receptor only or dual of CCR5/CXCR4 co-receptor. A BPDUCM seeking 10 possible partitions was applied to define V3 genetic determinants either independently or interactively associated with the usage of those three cases.

Efforts were also being taken to infer dependency structure among cooperative associations through RMS procedure. Positions 12, 28, 32 and 34 were further partitioned into two interaction subgroups 12 and 28 and 32 and 34.

The molecular basis underlying the mechanisms of action of such determinants has been further unraveled by docking analysis and free-energy calculations.

Specifically, our thermodynamic results confirmed the observation that mutations at positions 11, 12, 25, 28, 32 and 34 were found almost exclusively in CXCR4-using or dual-tropic viruses (Supplementary Table S50). In particular, G28E, I12M and I12T substitutions were found to be associated with remarkably favorable CXCR4 affinity profiles. Such a finding could be related to a strong electrostatic contribution of R9 for the co-receptor recognition in presence of the negatively charged glutamate that acts as a bridge between R9 and CXCR4 N-terminus (Supplementary Fig. S8). By contrast, in the I12M and I12T best poses, this weight decreased, as the co-receptor was able to establish one HB with C173 and good contacts with A394 (Supplementary Fig. S9).

Another interesting result was obtained in CXCR4 gp120 recognition in presence of the S11R substitution, associated with a favorable thermodynamic profile of the co-receptor. Specifically, in the best optimized configuration, we observed a strong electrostatic repulsion between the positively charged R11 and R9. Such a finding was related to a shift of the arginine at position 9 and its consequent involvement in an HB with a glutamic residue of the CXCR4 N-terminus, interaction not found in the WT sequence. Moreover in the WT gp120-CD4-CCR5 complex, serine at position 11 was involved in 3 HB with threonine at position 23; interestingly, in the WT gp120-CD4-CXCR4 optimized configuration, in addition to these hydrogen bonds, S11 established other 2 HB with N13. Consequently, we noticed a remarkable shift of the terminal part of V3 loop in gp120-CD4-CXCR4 complex, with a conformational fold responsible for a RMSD (Root Mean Square deviation), calculated on V3 residues 11-23 with respect to gp120-CD4-CCR5 complex, equal to 8.91 Å (Fig. 2).

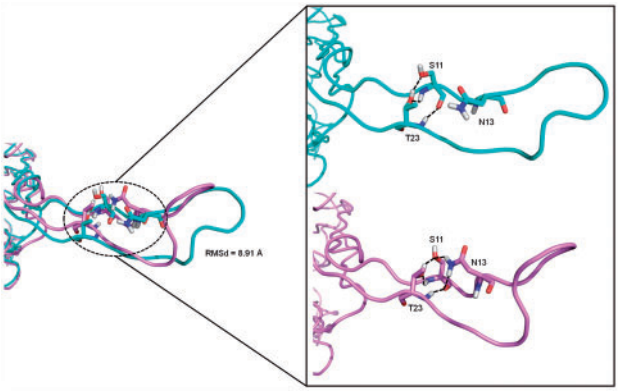

Fig. 2. Superimposition of gp120–CD4–CCR5 complex (cyan) with respect to gp120–CD4–CXCR4 complex (light magenta) in WT sequence. The glycoprotein is shown as ribbon, whereas the gp120 residues involved in HBs, indicated as black dashed lines, are represented as cyan and light magenta carbon sticks, respectively. The RMSd value is reported and expressed in Angstrom

Interestingly, position 34 was found to be related to the best CXCR4 affinity profile, in particular when mutation H34Y/N was associated with substitutions at position 32. Analyzing the component interaction data, such a finding could be justified by the strong contribution of C173 and S177 residues in CXCR4 recognition. In particular, the co-presence of mutations H34N + Q32K, associated with CXCR4 usage, was involved in the most significant affinity improvement because of the co-receptor ability in establishing two HB with C173. Moreover, the comparison of such a mutated complex with respect to the WT revealed some interesting structural insights. In fact, in the WT best pose CXCR4 was found able to establish two HB and several VdW interactions with Q32, whereas hystidine at position 34 resulted implicated in an intramolecular HB with N36. This contacts network could probably explain the conformational change observed in presence of the double-mutation H34N + Q32K. In particular in this case, no intramolecular interaction was noticed, and CXCR4 was found to get closer to the gp120 C2 conserved domain, not involved in the co-receptor recognition of the WT complex (Supplementary Fig. S10).

A previous study highlighted the presence of genetic determinants in C2 domain critical for co-receptor usage (even independently from the use of CD4) (Dumonceaux *et al.*, 2001). Here, we provide structural basis explaining the involvement of C2 domain in mechanisms underlying the usage of CXCR4 co-receptor.

Some interesting structural details were also revealed in presence of Q32K/R and H34Y mutations, with a particular remark when they were associated. Specifically, in Q32K and H34Y mutated complexes, CCR5 N-terminus was able to establish an HB with C173 and K175, respectively, by means of its tyrosine terminal residue. These interactions were probably responsible for the co-receptor shift from the gp120 V3 loop, with the

consequent loss of crucial contacts with E25 and R31 and a final unfavorable thermodynamic profile if compared with the WT sequence. Interestingly, the co-presence of Q32K and H34Y substitutions, associated with an increased CCR5-binding affinity with respect to the WT and to the single mutations, was related to a significant co-receptor conformational change because of a better accommodation of the sulfotyrosine at position 14, able to establish an additional HB with R3.

Conversely, when the glutamine at position 32 was replaced by an arginine, we observed a remarkable conformational change of the CCR5 sulfotyrosine at position 10. In particular, with respect to the WT complex, where Huang's geometries were well reproduced, Tys10 dihedral angle resulted completely inverted with respect to the WT, causing the lack of crucial HB with R31 (Supplementary Fig. S11). Moreover, CCR5 N-terminus terminal tyrosine was able to establish an additional HB with K175, with the co-receptor shift from the gp120 V3 loop and a final destabilizing effect in the glycoprotein recognition step.

In CXCR4 gp120 mutated complexes, we observed that the increased co-receptor affinity versus gp120 could be attributed to a conformational rearrangement of the sulfotyrosines. In particular, in presence of the single-mutations Q32K and Q32R, Tys21 resulted closer to V3 loop, with remarkable co-receptor stabilization because of its additional HB with S177. Interestingly, the presence of H34Y mutation, single or associated with the substitutions at position 32, caused a further conformational accommodation of CXCR4 N-terminus, with a shift of Tys7, which was well stabilized by several HB with N5, N6, N7 and T8 residues. Such an additional conformational change allowed us to assume that the remarkably increased CXCR4 affinity mainly depends on the mutation at position 34, rather than at position 32.

In conclusion, new genetic determinants of tropism within the V3 domain have been detected and confirmed by phenotypic and molecular modeling analysis. Additionally, our GBPM analysis led us to recognize the most relevant structural features of gp120 V3 loop associated with both co-receptors and the most pivotal protein–protein occurring interactions into all the analyzed complexes independently from their chemical nature. These observations could clear the way for the application of a virtual screening method able to identify novel scaffolds as potential entry-inhibitors and for the development of new pharmacophoric models helpful in the drug discovery, vaccine design and clinical disease management (Cardozo *et al.*, 2007; Parczewski *et al.*, 2010; Sander *et al.*, 2007).

Funding: J.Z. was supported by start-up funding and Sesseel Award from Yale University. The computation was done with the help from the Yale University Biomedical High Performance Computing Center, which was supported by the NIH (grant RR19895). This work was financially supported by grants from CHAIN, Collaborative HIV and Anti-HIV Drug Resistance Network, Integrated Project no. 223131, funded by the European Commission Framework 7 Program, and from the Italian Ministry of Health (CUP: E81J10000000001), and from Aviralia Foundation.

Conflict of Interest: none declared.

REFERENCES

Alcaro,S. *et al.* (2010a) Computational analysis of Human Immunodeficiency Virus (HIV) Type-1 reverse transcriptase crystallographic models based on significant conserved residues found in Highly Active Antiretroviral Therapy (HAART)-treated patients. *Curr. Med. Chem.*, **17**, 290–308.

Alcaro,S. *et al.* (2010b) Rational design, synthesis, biophysical and antiproliferative evaluation of fluorenone derivatives with DNA G-quadruplex binding properties. *ChemMedChem*, **5**, 575–583.

Berger,E.A. *et al.* (1998) A new classification for HIV-1. *Nature*, **391**, 240.

Bowers,K.J. *et al.* (2006) *SC '06 Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. New York, NY.

Cardozo,T. *et al.* (2007) Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res. Hum. Retroviruses*, **23**, 415–426.

Church,J.D. *et al.* (2008) HIV-1 tropism and survival in vertically infected Ugandan infants. *J. Infect. Dis.*, **197**, 1382–1388.

Dorr,P. *et al.* (2005) Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob. Agents Chemother.*, **49**, 4721–4732.

Dumonceaux,J. *et al.* (2001) Determination of essential amino acids involved in the CD4-independent tropism of the X4 human immunodeficiency virus type 1 m7NDK isolate: role of potential N glycosylations in the C2 and V3 regions of gp120. *J. Virol.*, **75**, 5425–5428.

Gasteiger,J. and Marsili,M. (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, **36**, 3219–3228.

Goodford,P.J. (1985) A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.*, **28**, 849–857.

Hoffman,T.L. and Doms,R.W. (1999) HIV-1 envelope determinants for cell tropism and chemokine receptor use. *Mol. Membr. Biol.*, **16**, 57–65.

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

Huang,C.C. *et al.* (2007a) Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*, **28**, 1930–1934.

Huang,W. *et al.* (2007b) Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous compositionof viral populations. *J. Virol.*, **81**, 7885–7893.

Jorgensen,W.L. *et al.* (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **118**, 11225–11236.

Kaminski,G.A. *et al.* (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474–6487.

Kuiken,C. et al. (eds) (2010) *HIV Sequence Compendium 2010*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM. LA-UR 10-03684.

Lihana,R.W. *et al.* (2009) HIV-1 subtype and viral tropism determination for evaluating antiretroviral therapy options: an analysis of archived Kenyan blood samples. *BMC Infect. Dis.*, **9**, 215–223.

McDonald,D.Q. and Still,W.C. (1992) AMBER torsional parameters for the peptide backbone. *Tetrahedron Lett.*, **33**, 7743–7746.

Mohamadi,F. *et al.* (1990) MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.*, **11**, 440–467.

Moreno,S. *et al.* (2009) Prevalence of CCR5-tropic HIV-1 among treatment-experienced individuals in Spain. *HIV Clin. Trials*, **10**, 394–402.

Ortuso,F. *et al.* (2006) GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics*, **22**, 1449–1455.

Parczewski,M. *et al.* (2010) Analysis of V3 loop sequences using various bioinformatic tools designed for genotypic HIV-1 tropism testing. *HIV AIDS Rev.*, **9**, 65–67.

Princen,K. and Schols,D. (2005) HIV chemokine receptor inhibitors as novel anti-HIV drugs. *Cytokine Growth Factor Rev.*, **16**, 659–677.

Regoes,R.R. and Bonhoeffer,S. (2005) The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol.*, **13**, 269–277.

Sander,O. *et al.* (2007) Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. *PLoS Comput. Biol.*, **3**, e58.

Shepherd,J.C. *et al.* (2008) Emergence and persistence of CXCR4-tropic HIV-1 in a population of men from the multicenter AIDS cohort study. *J. Infect. Dis.*, **198**, 1104–1112.

Still,W.C. *et al.* (1990) Semianalytical treatment of solvation for molecular mechanism and dynamics. *J. Am. Chem. Soc.*, **112**, 6127–6129.

Svicher,V. *et al.* (2011a) Identification and structural characterization of novel genetic elements in the HIV-1 V3 loop regulating coreceptor usage. *Antivir. Ther.*, **16**, 1035–1045.

Svicher,V. *et al.* (2011b) HIV-1 dual/mixed tropic isolates show different genetic and phenotypic characteristics and response to maraviroc *in vitro*. *Antiviral Res.*, **90**, 42–53.

Svicher,V. *et al.* (2011c) Key-genetic elements in HIV-1 gp120 V1, V2, and C4 domains tightly and differentially modulate gp120 interaction with the CCR5 and CXCR4 N-terminus and HIV-1 antigenic potential. *Antivir. Ther.*, **16**, A14–A14.

Svicher,V. *et al.* (2011d) Specific HBsAg genetic-determinants are associated with occult HBV-infection in vivo and HBsAg-detection. *Antivir. Ther.*, **16**, A85–A85.

Svicher,V. *et al.* (2012) Novel HBSag markers tightly correlate with occult HBV infection and strongly affect HBSag detection. *Antiviral Res.*, **93**, 86–93.

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

Veldkamp,C.T. *et al.* (2008) Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci. Signal.*, **1**, ra4.

von Itzstein,M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, **363**, 418–423.

Zhang,J. *et al.* (2010) Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *PNAS*, **107**, 1321.

Zhang,J. *et al.* (2012) Systematic investigation on interactions for HIV drug resistance and cross-resistance among protease inhibitors. *Proteome Sci. Comput. Biol.*, **1**, 2.