Published in final edited form as:

Curr HIV Res. 2014 ; 12(2): 132–161.

Bioinformatic Analysis of HIV-1 Entry and Pathogenesis

Benjamas Aiamkitsumrit, Will Dampier, Gregory Antell, Nina Rivera, Julio Martin-Garcia, Vanessa Pirrone, Michael R. Nonnemacher, and Brian Wigdahl*

Department of Microbiology and Immunology, and Center for Molecular Virology and Translational Neuroscience, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania 19102, USA

Abstract

The evolution of human immunodeficiency virus type 1 (HIV-1) with respect to co-receptor utilization has been shown to be relevant to HIV-1 pathogenesis and disease. The CCR5-utilizing (R5) virus has been shown to be important in the very early stages of transmission and highly prevalent during asymptomatic infection and chronic disease. In addition, the R5 virus has been proposed to be involved in neuroinvasion and central nervous system (CNS) disease. In contrast, the CXCR4-utilizing (X4) virus is more prevalent during the course of disease progression and concurrent with the loss of CD4⁺ T cells. The dual-tropic virus is able to utilize both co-receptors (CXCR4 and CCR5) and has been thought to represent an intermediate transitional virus that possesses properties of both X4 and R5 viruses that can be encountered at many stages of disease. The use of computational tools and bioinformatic approaches in the prediction of HIV-1 co-receptor usage has been growing in importance with respect to understanding HIV-1 pathogenesis and disease, developing diagnostic tools, and improving the efficacy of therapeutic strategies focused on blocking viral entry. Current strategies have enhanced the sensitivity, specificity, and reproducibility relative to the prediction of co-receptor use; however, these technologies need to be improved with respect to their efficient and accurate use across the HIV-1 subtypes. The most effective approach may center on the combined use of different algorithms involving sequences within and outside of the *env*-V3 loop. This review focuses on the HIV-1 entry process and on co-receptor utilization, including bioinformatic tools utilized in the prediction of co-receptor usage. It also provides novel preliminary analyses for enabling identification of linkages between amino acids in V3 with other components of the HIV-1 genome and demonstrates that these linkages are different between X4 and R5 viruses.

© 2014 Bentham Science Publishers

*Address correspondence to this author at the Department of Microbiology and Immunology, Drexel University College of Medicine, 245 N. 15th Street, Philadelphia, PA 19102. Telephone: 215-762-7598; Fax: 215-762-1955; bwigdahl@drexelmed.edu.

Send Orders for Reprints to reprints@benthamscience.net

CONFLICT OF INTEREST

The authors declare they have no conflict of interest.

PATIENT CONSENT

Declared none.

HUMAN/ANIMAL RIGHTS

Declared none.

Keywords

CCR5; coreceptor; CXCR4; HIV; mutual information; PSSM

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) was first identified as the causative agent of the acquired immunodeficiency syndrome (AIDS) in 1984 when a retrovirus was isolated from the peripheral blood of patients suffering from a progressive and often fatal disease involving the quantitative depletion of the CD4⁺ T-cell population. The disease involved an array of associated defects in many arms of the immune system as well as the nervous system as a result of the neuroinvasive properties of this pathogenic retrovirus [1]. Since then, many investigators have been drawn to study this virus in order to understand the immuno- and neuropathogenic mechanisms associated with HIV-1 replication in T cells, cells of the monocyte-macrophage lineage, and other cellular targets within several end organs including the brain. Many research discoveries have led to the development of new diagnostic tools, new therapeutic agents, and numerous observations that will pave the way to the development of the next generation of preventive and therapeutic agents as well as new vaccine strategies.

HIV-1 has been shown to be a retrovirus and a member of the lentivirus family. HIV-1 infection and disease has presented with many of the features typically associated with the lentiviral family, including a long period of clinical latency, persistent viral replication, and multiorgan infection, including the brain [2, 3]. Target cells infected by HIV-1 include cells within the immune system, primarily T cells and cells of the monocyte-macrophage lineage (including dendritic cells) and cells within the brain, including endothelial cells, perivascular macrophages, astrocytes, and microglial cells [4, 5]. Viral entry into most cells has been shown to be initiated by attachment of viral gp120 to the host cell receptor, the CD4 molecule, and subsequently by viral binding to host cell chemokine co-receptor proteins (Fig. 1). While many co-receptors have been identified to play a role in HIV-1 entry, CXCR4 and CCR5 are the major co-receptors that have been characterized as supporting HIV-1 entry *in vivo* [6]. Upon binding, HIV-1 gp120 undergoes structural rearrangements involving conformational changes that lead to a change in HIV-1 gp41 (along with gp120, a cleavage product of gp160) from a nonfusogenic to a fusogenic state. This change brings the cellular membrane and viral envelope into closer proximity, thereby facilitating membrane fusion between the virus and target cell [7]. Subsequently, the viral core enters the host cell cytoplasm, and during this entire process the viral enzyme reverse transcriptase initiates the conversion of viral genomic RNA into a double-stranded DNA proviral genome. The proviral genome is then imported into the nucleus and integrated into the host cell genome by viral-encoded integrase [8, 9]. Subsequently, the viral promoter or long terminal repeat (LTR) directs transcription of the viral genome from a chromatin-based microenvironment [10–14]. Once the viral protein Tat accumulates by translation from a small pool of long cytoplasmic RNA transcripts, the production of full-length transcripts is greatly enhanced, fueling the replication process and production of high levels of infectious virus, particularly in the activated CD4⁺ T-cell population [10, 15–17]. This review focuses on the viral

envelope and cellular proteins (receptors and co-receptors) involved in the entry step; viral tropism for particular cell populations during the course of HIV-1 disease; and the utility of *in silico* co-receptor prediction methods and bioinformatic tools to determine co-receptor usage by HIV-1.

OVERVIEW OF THE HIV-1 ENTRY PROCESS

The entry of HIV-1 into target cell populations is a receptor-mediated, pH-independent process based on the direct interaction between the viral-encoded gp120 and a host cell receptor molecule (CD4) as well as one of the co-receptor molecules, the most well characterized and prominent of which are CXCR4 and CCR5 [18–20]. The CD4 molecule is a 60-kDa glycoprotein expressed at different levels on the surface of lymphocytes, cells of monocyte-macrophage lineage, and cells within the CNS, including perivascular macrophages and microglial cells [21]. The most well-known role for CD4 within the immune system is in signaling between T and B lymphocytes as well as in providing an antigen-induced activation of T-helper cells [22] and modulating CD8⁺ T-cell functions [23]. In addition to these normal cellular functions, in 1984, the CD4 molecule was shown to serve as the primary cellular receptor for HIV-1 entry [24–27]. A series of monoclonal antibodies directed against the CD4 molecule were shown to block syncytia formation and inhibit the production of vesicular stomatitis virus pseudotyped with the HIV-1 envelope in selected susceptible cell types [25]. Furthermore, preincubation of CD4⁺ T cells with three different antibodies directed against different epitopes of the CD4 molecule was shown to block HIV-1 infection *in vitro* [26].

The interaction between viral gp120 and the CD4 molecule has been shown to promote the association of the gp120-CD4 complex with another membrane component, the co-receptor (Fig. 1). A recent study utilizing small-angle X-ray scattering and hydrogen/deuterium exchange technologies confirmed that an unliganded full-length gp120 was actually dynamic and revealed the V1/V2 loops in proximity on the top of the molecule [28]. Once gp120 binds to the CD4 molecule, the V1/V2 region, which has already been shown to be in contact with the V3 loops, subsequently un.masks the neighboring co-receptor binding sites, thereby rearranging and changing the orientation of the bridging sheet formation in the gp41-interactive stage [28]. The gp120-CD4 interaction was also demonstrated by experiments showing an increase in specific antibody binding to specific regions of gp120 and gp41 when HIV-1-infected cells form complexes with soluble CD4 [29]. A physical association of a receptor known as fusin and the gp120-CD4 complex on the cell membrane was shown, suggesting that this association likely contributed to the exposure of the gp41 hydrophobic NH₂-terminus [30]. Fusin was later determined to be the chemokine co-receptor molecule CXCR4 [31], and this protein was identified as the co-receptor for CXCR4-utilizing HIV-1 isolates [32]. The chemokine receptor CCR5 was subsequently discovered as an additional co-receptor for a substantial number of HIV-1 strains, and these viral isolates were designated CCR5-utilizing (R5) viruses [33–37]. Other chemokine receptors have also been demonstrated to support HIV-1 infection [34, 38, 39]. The seven-transmembrane co-receptors, Apj and CCR9, discovered in 1998, support the efficient entry of several primary CXCR4-utilizing and dual-tropic HIV-1 isolates, which can utilize both CXCR4 and CCR5 chemokine co-receptors for viral entry [39]. In addition, the β chemokine

G protein–coupled receptor (GPCR) CCR3 was reported to facilitate HIV-1 infection by primary isolates of HIV-1 [34]. The β chemokine receptor CCR-2b (along with the fusin co-factor) has also been shown to serve as a co-receptor for some macrophage-tropic (M-tropic) strains of virus, as well as for the dual-tropic 89.6 strain [37].

THE HIV-1 ENVELOPE GENE AND ITS TRANSLATIONAL CLEAVAGE PRODUCTS GP120 AND GP41

The HIV-1 envelope gene (*env*) is 2,568 bp in length and its product is termed gp160. The envelope gene consists of 1,449 bp encoding gp120 and 1,035 bp encoding gp41 (www.lanl.gov). HIV-1 gp160 is the initial precursor protein synthesized in the endoplasmic reticulum as a single polypeptide precursor, glycosylated, and assembled into oligomers [40–43]. The cleavage of the precursor gp160 into the noncovalently associated gp120 and gp41 occurs in the Golgi apparatus [44] by cellular proteases such as furin and precursor convertases (PC) 5/6 and PC6, which have been suggested to play a major role in gp160 cleavage in CD4⁺ T lymphocytes [45]. The gp120 subunit is then delivered to the envelope of nascent viral progeny [46].

Glycoprotein 120 (gp120)

The HIV-1 gp120 subunit is comprised of a 60,000 Da core polypeptide that contains highly N-linked glycosylation modifications, previously shown to increase the molecular weight up to 120,000 Da [47]. The gp120 molecule contains five variable regions (V1-5) that alternate between five constant regions (C1-5) [42]. The amino acid sequences of the variable regions are highly modified by insertions, deletions, and substitutions [42]. Regardless of the hypervariable sequences, the overall structural and functional elements of gp120 are relatively well conserved [42].

The first and second variable loops (V1 and V2) represent the largest part of gp120 and play an important role in envelope assembly and function, viral entry, antibody-induced neutralization, CD4 binding, co-receptor usage, and viral spread [48–53]. Despite being hypervariable, V1 and V2 contain a cluster of relatively conserved residues, among which are nine charged amino acids, a greater number than are present in other regions of gp120 [54]. In addition, within these nine charged amino acids, the aspartic acid at position 180 has been considered highly conserved among HIV-1 isolates [54] and is crucial in the early steps of viral entry, as confirmed by alanine substitution studies [55]. Mutagenesis of the aspartic acid at position 180 to alanine results in an alteration in viral infectivity and slower viral growth kinetics. However, the incorporation of gp120 and gp41 into the viral particle, and their ability to subsequently bind CD4, were shown not to be affected by this sequence alteration [55]. Other studies have shown that deletion of either V1 or V2 alone had a minimal effect on envelope function and viral replication in a cell type–specific manner [56]. However, deletion of both the V1 and V2 loops impaired envelope function, including impaired viral entry [57, 58] and replication in primary CD4⁺ T cells [59]. This was explained by the absence of an $\alpha 4\beta 7$ recognition motif, which is located within the V2 loop region. The lack of this motif resulted in the downregulation of expression of the LFA-1 adhesion molecule in infected cells, resulting in decreased viral spread. However, an

alternative explanation of these results has suggested that the deletion of the V1/V2 region altered the interaction of the viral particle with CD4 and/or CXCR4 [59].

In conjunction with previous observations, it has been reported that alteration of the V1/V2 loops can modulate the viral phenotype with regard to co-receptor usage, CD4 binding, antibody neutralization, and its association with changes in HIV-1 disease severity [60, 61]. A mutated primary R5 HIV-1 isolate, SF162 V1 V2, which has been shown to lack the V1 and V2 loops, exhibited an overall normal CD4 binding site conformation on viral surfaces because the binding pattern of antibody directed against the CD4 molecule was similar to that of cells infected with the parental virus [56]. Nevertheless the antibody recognition profiles were different between V1-negative virus, V2-negative virus, and the parental virus, SF162, suggesting a different structure and exposure of epitopes located on the viral gp120 molecule [56]. These V1- and V2-mutant viruses exhibited a comparable level of replication within peripheral blood mononuclear cells (PBMCs) as compared with the wild-type parental virus [56]. The decrease in viral replication in macrophages has been observed with V1- or V2-deleted viruses, and it is therefore correlated with a post-entry step because the CD4 binding and co-receptor usage profile were not altered in these viruses, as compared with the parental virus, further indicating that the role of V1 and V2 in viral entry is cell type-specific [56]. The neutralizing antibody susceptibility profiles of V1-deleted SF162 (SF162 V1), V2-deleted SF162 (SF162 V2), V1/V2-deleted SF162 (SF162 V1 V2), and parental SF162 were shown to be different. Removal of the V2 loop resulted in increased neutralization susceptibility to most of the antibodies examined, which might be because the overall V2 region was involved in a number of interactions with different gp120 epitopes [51]. The susceptibility of SF162 V2 and SF162 V1 V2 to antibody against the CD4 binding site was increased potentially due to the CD4 binding site of these viruses being more exposed based on the lack of the V2 loop region [51]. In contrast, SF162 V1 exhibited resistance to neutralization by antibody against CD4i epitopes, CD4 binding site, and V3 loop, potentially due to readjustment of the gp120 glycoprotein resulting in decreased epitope exposure, however, with no disruption in the ability of the virus to enter and replicate [51]. Another study showed that HIV-1 lacking the V1 and V2 regions exhibited approximately 30% less entry capability in Jurkat T cells; was syncytium-forming incompetent; exhibited delayed replication kinetics; and had increased sensitivity to neutralization antibodies as compared with the wild-type parental virus [60]. A recent vaccine efficacy study focused on preventing HIV-1 infection also showed that amino acid positions 169 and 181 within the V1/V2 region were involved in elevating the immune response to HIV-1, suggesting that specific genetic signatures within the V1/V2 region could contribute to the exposure of epitopes targeted by neutralizing antibodies [62]. Additionally, a molecular structure study has proposed a model in which V1/V2 functions as a shield for the V3 region because deletion of V1/V2 regions leads to elevation of neutralizing antibodies directed against V3 epitopes [63].

In summary, V1 and V2 serve as concealers of the CD4 binding site; thus, deletion of both of these regions results in increasing the sensitivity to neutralization antibodies. Deletion of either V1 or V2 alone appears to display opposite effects in defining HIV-1 neutralization susceptibility; deletion of the V2 region enhances the exposure of epitopes on gp120 whereas deletion of only the V1 region has no effect on the level of neutralizing antibody

sensitivity. Overall, viral gp120 lacking operational V1 and/or V2 loops affects viral infectivity and replication potential in a cell type-specific manner [60, 64].

The variable loop region 3 (V3) has been intensively studied because it was identified as the principal neutralizing domain (PND) on the viral envelope glycoprotein [65, 66] and was considered to be hypervariable with regard to amino acid sequence [67]. The V3 loop consists of approximately 34 to 36 amino acid residues, at positions 296 to 331 of the HIV-1 HXB2 strain [68]. Based on nuclear magnetic resonance (NMR) analysis, V3-derived peptides exhibit two distinctive sequences and conformations, which are similar to two different groups of natural ligands of CCR5 and CXCR4. Based on this structural information, it has been suggested that the V3 loop was responsible for the selective interaction with the two different co-receptors (CXCR4 and CCR5) [69] (Fig. 2).

Several studies of V3 have focused on its role in viral entry and co-receptor usage. Several investigators have examined V3 loop sequence variation with respect to viral phenotype. In this regard, two major approaches have been taken to analysis of the observed variations. First, studies were performed to determine the covariant sequence alterations within the V3 loops themselves, and second, the sequences were compared to a specific strain or consensus sequence [68]. Analysis of the covariant sequence alterations within the V3 loop has focused on determining the mutual information between amino acid residues. The most commonly encountered covariant residues detected when 308 distinctive V3 loop sequences were aligned consisted of the residue pairs of 24/25, 13/25, 13/19, 13/24, 20/25, 11/25, and 11/13 [70]. However, the amino acid positions 11 and 25 were shown to be consistent with the charge rule, and this pair of residues has been widely used for predicting viral phenotypes [71]. When the V3 loop region of HXB2 was replaced with isolates that exhibited nonsyncytium-inducing (NSI) and syncytium-inducing (SI) phenotypes from the same patient, the NSI virus displayed a serine (S) at position 11 and aspartic acid (D) at position 25; however, arginine (R) at position 11 and glutamine (Q) at position 25 were found in the SI viruses [71]. As previously described, the charge rule has been applied to the total net charge of V3 sequences, which varies globally between +2 to +10 [72]. NSI viruses exhibited less positive charge than SI viruses [73], and similar results were demonstrated when comparing M-tropic versus non-M-tropic viral isolates [73].

The viral phenotype transition has been shown to relate to co-receptor usage. In this regard, a T-cell tropic (T-tropic) isolate or SI virus was discovered to use primarily CXCR4 as a co-receptor for viral entry. In contrast, the primary NSI isolate-exhibiting phenotype was shown to preferentially use CCR5 as the entry co-receptor [32, 35]. The 11/25 and net charge rules have previously been used to identify the viral phenotype in the context of co-receptor selection, SI capability, and tropism. These studies have demonstrated that if a positively charged amino acid was located at position 11 and/or 25 within the V3 loop and the total net charge resided between 5 and 8, the virus would exhibit CXCR4 usage, or an X4 virus and SI phenotype [71, 74]. Conversely, if the amino acid at position 11 and/or 25 was neutral or negatively charged and was accompanied by a total net charge that fell between 2 and 4, the virus would exhibit CCR5 usage, or an R5 virus with an NSI phenotype [73–75]. Dual-tropic virus (X4/R5) has been defined by its ability to use either CXCR4 or CCR5 as a

co-receptor, and these viruses generally score like an X4 virus based on the application of both the 11/25 and net charge rules [74].

The strategy of amino acid charge and co-receptor usage in phenotyping viral isolates has been supported by protein crystal structures and NMR studies that have indicated that amino acid positions 11 and 25 were located at the end of the V3 crown [72] and are found opposite of each other with their charges capable of leading to an electrostatic interaction that can bind chemokine ligand-like structures [76, 77]. Amino acid residues 1 to 10 and 26 to 35 were designated as the N- and C-termini of the V3 stem, respectively. Residues 11 to 25 were designated as the V3 crown [72]. The V3 stem and crown have distinct functional domains [68]. The V3 stem alone, for example, has been shown to mediate soluble gp120 binding to CCR5 aminoterminal domain-based sulfopeptides; however, both the stem and crown are necessary for binding of soluble gp120 to cell-associated CCR5 [72].

Nevertheless, within the context of the virion, the V3 crown alone was sufficient for determination of co-receptor usage [72]. The charges of amino acid residues 11/25, which are part of the crown, are likely involved in the selection of the co-receptor *via* electrostatic interaction. The extracellular loops (ECLs) of CXCR4 have been shown to contain higher negative charge than that of CCR5 [78]. It has been suggested that the higher total positive charge of an X4 V3 would interact with CXCR4 better than with CCR5. Even though the V3 loops are highly variable, the GPG motif located in the middle of the PND, within the crown, has been shown to be well conserved among HIV-1 isolates [67]. The changes that occur around this GPG motif alter the structure of the V3 crown and/or surface accessibility and thereby may guide co-receptor selection [68].

Variable loop 4 (V4) of HIV-1 is approximately 19 to 26 amino acid residues long and is overall negatively charged [79]. According to the sequence alignment, the V4 loop has been shown to be as variable as V1 in all HIV-1 subtypes, except subtype E [79]. In studies using the simian immunodeficiency virus (SIV)-infected macaque model, neutralizing epitopes recognizing the V4 region were detected, and changes in the amino acid profile within V4, especially at positions 412 to 418, were shown to alter the N-linked glycosylation pattern, resulting in alteration of antibody recognition and an overall neutralizing profile [80]. In addition, the amino acid mutation found within V4 resulted in the production of a new site for glycan modification, and this alteration was shown to play a role in disease progression in simian AIDS [81]. However, in HIV-1 disease, neutralizing antibody against V4 has not yet been described [79]. This may be due to the frequency of glycosylation modification detected within V4. The potential N-linked glycosylation found in V4 in SIV is less than that of the HIV-1 V4; consequently, less carbohydrate masking has been observed during the course of SIV infection, thereby resulting in a greater response to neutralization antibody as compared with HIV-1 [79, 81, 82]. Given this observation, neutralizing antibody against V4 has not been found in HIV-1 infection in humans; the core motif of the superantigen (SAg)-binding site has been identified on a discontinuous epitope spanning V4 [83]. The SAg can engage B cells and activate VH3 B cells *in vitro* [84, 85]. Mutational analysis of V4 demonstrated significant binding of defective SAg, suggesting a role for V4 in gp120 SAg binding to normal human immunoglobulins [83]. Furthermore, the binding of SAg and VH3 B cells to the V4 region may directly explain the dysfunction and depletion of B cells

during disease progression. The potential sites for glycosylation on HIV-1 gp120 have been identified [42], and based on the high degree of glycosylation, the virus has been shown to exhibit altered receptor binding sites and neutralization recognition [86]. Within the V4 loop, amino acid position 385 is involved in the formation of a disulfide bond with amino acid position 418 and the glycan at position 386, which is located at the base of the V4 loop, demonstrating their importance in neutralization sensitivity by which their removal resulted in resistance to the 2G12 antibody [87]. N386 does not have an impact on envelope folding, but N-linked glycosylation has been shown to facilitate immune evasion by protecting the CD4 binding site against antibodies [87]. In addition, removal of this glycan by mutating amino acid N386K has been shown to enhance viral infection, and this process is CD4 independent [88]. Therefore, removal of the glycan at N386 may enhance therapeutic strategies by increasing neutralizing antibody against the CD4 binding site and the SAg motif on V4 and may serve as a potential candidate vaccine target to elicit antibody production.

Variable loop 5 (V5) of HIV-1 gp120 has been shown to contain approximately 7 to 11 amino acid residues, occasionally found to be 11 to 15 amino acids in length [89], with an average of two N-linked glycosylation sites [79]. Very little functional information has been obtained concerning V5. Sequence analysis of the HIV-1 gp120 V5 region has shown this loop structure to be hydrophilic with a net negative charge [79]. Sequence variation or mutations at the glycosylation sites within V5 enhance viral escape from neutralizing antibodies, which result from the alteration of gp120 envelope structure and epitope exposure [90–92]. One study has shown that removal of a V5 N-linked glycosylation site by mutagenesis at N462 led to resistance to anti-gp41 antibodies [93]. The level of genetic diversity studied within V5 demonstrated a small increase during the first year of infection in patients undergoing antiretroviral therapy (ART), and a higher level of genetic diversity was observed in patients naïve to ART [89]. With respect to the production of HIV-1 gp120 V5 region neutralizing antibody during human infection, antibodies directed at the V5 region have not been reported within an HIV-1-infected hemophiliac cohort [79]. However, in a recent vaccine study, neutralizing activity against V5 was identified in sera from rabbits that were immunized with the HIV-1 gp120 subunit derived from the JR-CSF envelope [94].

Glycoprotein 41 (gp41)

HIV-1 gp41 has been shown to be comprised of approximately 345 amino acids and organized into three major regions including the ectodomain, membrane-spanning domain (MSD), and cytoplasmic tail [79]. The ectodomain consists of an extended, triple-stranded alpha-helical coiled coil with an exposed aminoterminal at the tip. The carboxyterminus consists of an alpha helix lying in an antiparallel direction against the outside coil [95]. Therefore, the N- and C-termini have been located at the opposite ends of the ectodomain, suggesting that the ectodomain is a functional component in the initial events associated with membrane fusion [95].

The ectodomain consists of the fusion peptide as well as leucine zipper-like (LZL) and assembly domains. All of these domains are important for the overall structure and function of the envelope glycoprotein [79]. A hydrophobic fusion peptide has been identified at the

N-terminus of gp41 and contains a number of Phe-Leu-Gly motifs [96]. Introducing amino acid mutations within the fusion peptide region demonstrated that substitution of hydrophobic amino acids with polar amino acids blocked or reduced syncytium formation capability but did not affect the binding of the viral envelope to the CD4 molecule [97]. The LZL motif has been shown to be conserved among HIV-1 isolates and contain a 4–3 heptad repeat of hydrophobic residues [98, 99]. It was suggested that most of the leucine zippers mediate dimerization of several transcription factors [100, 101] and this leucine zipper motif may also be involved in dimerization of the HIV envelope glycoproteins [102]. In order to understand the leucine zipper motif completely, mutational analyses have been performed [103]. Deletion of 12 amino acid residues at the N-terminus of the gp41 alpha-helix/leucine zipper resulted in interference with gp41 oligomerization but did not completely affect the oligomerization of gp160 [103]. In a second study, deletion of 12 amino acid residues at the C-terminal domain of the gp41 leucine zipper motif did not affect the oligomeric structure of gp41 and gp160; however, a combination of the 12 amino acid residue deletion at the C-terminus with the previously described 12 amino acid deletion at the N-terminal domain caused a total disruption of both gp41 and gp160 oligomeric structures [103]. Finally, deletion of 11 amino acid residues adjacent to the membrane-spanning region of gp41 resulted in a partial effect on the oligomeric structure of gp41 but had no effect on the structure of gp160 [103]. Therefore, it was suggested that amino acid residues at the N-terminal domain of the gp41 alpha-helix/leucine zipper were not only essential for the maintenance of the oligomeric structure of gp41, but together with amino acid residues at the C-terminal domain were required to mediate the oligomerization of HIV-1 envelope glycoprotein 160 [103].

The MSD of gp41 consists of approximately 25 amino acids and is highly conserved among the viral isolates examined to date, especially the core region (K681 to R694), which is buried within the lipid bilayer [104]. The MSD is involved primarily in anchoring the envelope glycoprotein complex to the lipid bilayer [105]. Deletion of the MSD, through introduction of stop codons and addition of polar amino acids, resulted in a reduction of cell-associated envelope protein and increased levels of soluble envelope protein found in the supernatant of transfected cells [106]. In addition, alterations in the MSD region caused different degrees of posttranslational proteolytic processing of gp160 as compared with posttranslational processing of wild-type parental MSD [106].

It was demonstrated that all 25 amino acids of the gp41 MSD were not required for biological function of Env; the 17 amino acids in the core region were sufficient to support the anchoring of gp41 in the membrane and to mediate cell-to-cell fusion capability of the virus [107, 108]. The amino acid sequence of the lentiviral MSD differs from most other viral envelope proteins based on the presence of charged amino acids near or within the MSD in comparison to the stretches of hydrophobic amino acids of most other viral envelope structures in this region [109, 110]. This unusual structural feature of the HIV-1 MSD is necessary for Env-mediated fusion [105, 111]. As previously shown, the changes in hydrophobicity of the region proximal to the gp41 MSD affect syncytium formation and virus replication, despite the fact that neither the envelope glycoprotein expression on the surface of the infected cell or the surface of the virus particles nor the ability to bind to the CD4 molecule has been altered [111]. The amino acid residues from positions 696 to 707 of

the transmembrane domain have been shown to exhibit a functional importance in fusion; analyses of point mutations at amino acid position 696 alone or in combination with 707 and 709 or 698 with 707 and 709 have resulted in deficient syncytium formation [111]. Truncations of amino acid positions 695 to 697, positions 698 to 706, and positions 695 to 706 have also shown deficiency with respect to fusogenicity [111]. However, the cell surface expression of these MSD mutants was not altered [111]. Investigators have suggested that basic amino acid residues of the transmembrane region between position 696 and 707 are crucial for fusogenicity but not for cell surface transportation or anchoring [111].

The cytoplasmic tails of HIV-1 and HIV-2 gp41 are approximately 150 amino acids in length and are considered unusually long, compared with those of other retroviral envelopes. It has been demonstrated in a cell type-specific manner that the long HIV-1 gp41 cytoplasmic tail is essential for stability of the envelope glycoprotein [112] and/or for incorporation of the envelope glycoprotein into mature virions [113, 114]. Deletion of 17 amino acid residues at the C-terminus of the cytoplasmic domain resulted in decreased stability of the envelope glycoprotein in infected COS-1 cells and in Jurkat T cells. Interestingly, incorporation of the envelope into the virion was not affected in either cell line examined [112]. Conversely, HIV-1 virions with truncated gp41 demonstrated a remarkably low level of envelope glycoprotein on the cell surface in nonpermissive T-cell lines and monocyte-derived macrophages (MDMs). However, the expression of gp160 was comparable to the wild-type virus [113]. Deletion of 104 amino acid residues from the C-terminus of gp41 did not have any impact on the association of the envelope with the cell surface; however, truncation of 147 amino acids from the C-terminus of gp41 significantly reduced the stability and cell surface expression of the envelope glycoprotein in transfected COS-7 cells [114]. Additionally, mutations in the cytoplasmic domain of gp41 have demonstrated the impairment of viral infectivity in a cell type-dependent manner. Large truncated gp41 mutant virus was able to replicate in MT-4 cells, but not in a majority of T-cell lines, primary PBMCs, or MDMs [113]. Introduction of small deletions of the cytoplasmic tail demonstrated different effects on viral infectivity in COS-1 cells [115]. Deletion of 76 amino acid residues at the C-terminal did not impair viral infection; however, deletion of amino acid residues 6 to 14 altered the viral replication pattern, infectivity, and cytopathogenicity in COS-1 cells [115]. In addition, truncation of 17 amino acids from the cytoplasmic tail in a mutant virus exhibited reduced replication in Jurkat T cells, but it exhibited similar levels of viral syncytia formation and cell viability when it reached a similar level of virus production [112]. Truncation of the cytoplasmic domain from amino acid residues 6 to 192 also resulted in failure of virus to infect both H9 and CEMx174 T-cell lines [116]. Conversely, truncation of 144 amino acid residues in the cytoplasmic domain exhibited slightly delayed infection kinetics compared with wild-type virus [117]. Therefore, the cytoplasmic domain of gp41 affects viral infectivity, determined by not only altering the length, specific sequences, and/or conformational aspects of the gp41 cytoplasmic tail, but also by the cell type-specific nature of the mutant phenotype.

In conclusion, HIV-1 gp120 is a surface protein that is responsible for binding receptor and co-receptor molecules, thereby facilitating viral entry. Each step of binding has been associated with structural rearrangements leading to greater exposure of viral epitopes,

which become more recognizable by neutralizing antibodies. HIV-1 gp120 is highly variable, containing a number of domains with extensive variation with many regions of the protein altered by covalent modifications that facilitate immune evasion. Regardless of hypervariability, the overall structure of gp120 is fairly well conserved. The V1/V2 loop region is critical for neutralization antibody evasion and functions as a region that is involved in masking epitopes within the V3 region. The V3 loop has been studied extensively because of its major role in co-receptor binding and differential use of CXCR4 and CCR5 by X4 and R5 viruses, respectively. By contrast, few studies have focused on the V4 and V5 loops. Instead, additional studies have concentrated on the region of gp120 involved in CD4 binding, which has been located between the V4 and V5 regions. HIV-1 gp41 is necessary for anchoring the viral gp120 on the cell surface and in the process of membrane fusion. HIV-1 gp41 has been shown to contain charged amino acid residues within the MSD and has an unusually long C-terminal tail. Mutations within many regions of gp41 result in the loss of not only viral fusogenicity but also the stability of gp120 on the viral envelope surface.

CO-RECEPTORS

The host cell CD4 molecule represents the primary receptor for viral gp120 binding during the viral entry step. However, the CD4 molecule alone is not efficient for fusion and entry to occur; additional co-receptor molecules on the cellular surface also play critical roles in efficient viral entry [21, 118, 119]. Several studies have shown that different strains of HIV-1 are able to use different co-receptors preferentially, with implications with respect to viral phenotype, cellular tropism, viral transmission, and progression of HIV-1-induced disease in the peripheral blood and end organs such as the brain [120]. Many co-receptors used by HIV-1 have been discovered, including β -chemokine receptors CCR-3 [37], CCR-2b [37], CCR9 [39], Apj [39], and CCR8 [121]. However, the most extensive studies have focused on the chemokine receptors CXCR4 and CCR5.

Chemokine receptors belong to the transmembrane GPCRs, which have been shown to be comprised of seven-transmembrane helices [119, 122]. Interaction between specific ligands and the corresponding GPCR triggers a signal through the heterotrimeric GTP-binding protein and is involved in many acute and chronic inflammatory diseases, including HIV-1 infection [123]. The chemokine family has been shown to be comprised of a number of small cytokines that contain four conserved cysteine residues linked by disulfide bonds [123]. CXC and CC are the subfamilies of chemokines that have been classified by the position of the first two cysteine residues in the N-terminal domain and the other cysteine residue in ECL3, whether they are separated by other kinds of amino acids, and their respective receptors have also been studied extensively [123]. These two cysteine residues have been thought to be important in ligand-receptor binding conformation [123].

CXCR4

CXCR4 or CXC chemokine receptor 4, previously named fusin, was first identified as an orphan receptor and later was discovered to be an HIV-1 entry co-factor using a cDNA cloning strategy [32]. CXCR4 is composed of 352 amino acids and has been divided into an N-terminus, seven-transmembrane helices that are separated into three ECLs and three

intracellular loops, and a C-terminal domain; however, it exhibits distinct similarity to other chemokine receptors as shown by the fact that 32% to 38% of the amino acids present in these receptors are identical [123, 124]. CXCR4 has been shown to be expressed by a wide variety of cells, such as immune cells [125], cells within the CNS [126], and some cancer cell types [127]. The natural ligand for CXCR4 is stromal cell–derived factor-1 (SDF-1) or CXCL12. CXCR4 has 21 potential phosphorylation sites within the C-tail that are rapidly phosphorylated during cellular activation [128]. The activation of CXCR4 induces several downstream signaling pathways including the activation of protein kinase C, phosphatidylinositol 3 kinase, and extracellular signal-regulated kinases [129, 130]. These activation endpoints are involved in activation of transcription factors, promote cell growth and malignancy [131], and stimulate cell migration and lymphocyte chemotaxis in response to chemokine stimulation [132]. Deletion of CXCR4 or CXCL12 in mice has resulted in embryonic lethality, vascular and CNS developmental defects, impaired hematopoiesis, and cardiogenesis [124, 133, 134].

The role of CXCR4 as a co-receptor in HIV-1 entry was demonstrated by both gain-of-function and loss-of-function assays. Nonhuman NIH 3T3c cells expressing both CD4 and CXCR4 permitted HIV-1 envelope-mediated fusion and infection. In contrast, expression of anti-CXCR4 antibodies was able to potently block HIV-1 infection and fusion of CD4⁺ T lymphocytes [32]. An NMR structure of CXCR4 implicated the sulfotyrosine-containing N-terminus in binding to HIV-1 gp120, which has been shown to be analogous to CXCL12 recognition [124]. In support of this, U-373-CD4-LTR *LacZ* cells, engineered to express a series of deletion mutations in the N-terminal extracellular domain of CXCR4, three different T-cell line adapted viruses (LAI, RF, GUN-1), including a dual-tropic strain virus (89.6) exhibited different levels of impairment with respect to ability to fuse, enter, and replicate within specific target cells [135]. An additional study has shown that deletion of part of the N-terminal domain from amino acid positions 4 to 36 resulted in a 75% reduction in cell surface CXCR4 expression, a reduction of more than 50% in HIV-1 LAI infection (subtype B), and complete inhibition of HIV-1 strain NDK infection (subtype D) [136, 137]. Introduction of mutations involving changes in a tyrosine (Y) to alanine (A) at positions 7 and 12 of the N-terminal domain resulted in a decrease in efficiency of HIV-1 strains LAI and NDK infections, with both viruses exhibiting only 25% infection as compared with the parental CXCR4 [137]. These results indicated that the CXCR4 N-terminal domain played an important role in HIV-1 fusion and entry. In addition, different strains of HIV-1 exhibited different patterns of interaction with CXCR4 during infection [135–137].

The ECL as well as transmembrane regions of CXCR4 appear to be necessary for HIV-1 co-receptor activity [119]. A point mutation at amino acid position 193 (D193A), which was located in the ECL2 region; position 262 (D262A), which was located in the ECL3 region; or position 97 (D97N), which was located in the transmembrane region, caused both HIV-1 LAI and NDK infections to abate, suggesting that these amino acids were involved in electrostatic interactions between CXCR4 and viral gp120 [137].

Increased levels of SDF-1, the natural ligand for CXCR4, inhibited T-cell line tropic virus infection [138], and have resulted in a slower progression of disease in HIV-1-infected individuals who have been shown to carry specific polymorphisms in the SDF-1 chemokine

gene [139]. Experimental studies using CXCR4 chimeras and site-directed mutagenesis revealed that the N-terminal domain is necessary for SDF-1 binding; however, in order to trigger the downstream signaling, ECL2 is also required [140]. Studies have also shown that CXCR4 containing either D20A or Y21A in the N-terminal domain or E268A in ECL3 mutations decreased SDF-1 binding activities by more than 50% as compared with wild-type co-receptor; however, mutation in ECL2 at position 190 (Y190A) caused SDF-1-mediated signaling impairment [141]. In order to activate SDF-1-mediated signaling of CXCR4, the Asp-Arg-Tyr motif in the ECL2 was also required, but the C-terminal domain was not required [140]. Serial mutagenesis within the transmembrane region of CXCR4 has also shown that selective amino acid residues are important in HIV-1 cell fusion and infection, but these residues do not affect SDF-1 binding [142]. Therefore, both SDF-1 and HIV-1 required different regions of CXCR4 in order to bind and mediate downstream effects, and these regions overlap [140–142]. The CXCR4 mutants that fail to bind or mediate SDF-1 signaling were able to sustain viral infections. These observations suggest that CXCR4 can serve two different independent functions that involve binding to the natural receptor ligand and operating as a co-receptor for HIV-1 infection [140, 142]. As with an HIV-1 co-receptor, the N-terminal domain, ECL2, and ECL3 of CXCR4 have been shown to be involved in HIV-1 entry [140–142]. In addition, ECL2 has been shown to be restricted to particular strains of HIV-1 such as T-tropic and dual-tropic viruses [143]. Altering the charge of the amino acid located at position 187, which has been located within ECL2, from aspartic acid to other noncharged amino acid residues caused a reduction in infectivity of the T-tropic virus strain IIIB, as well as the dual-tropic strain 89.6. In addition, this mutation increased the infectivity of M-tropic viruses such as ADA, BaL, and JRFL strains [143].

CCR5

CCR5 has been another chemokine receptor that has been shown to belong to the GPCR superfamily located on chromosome 3 [123, 144]. It was first discovered in 1996 as ChemR13, by cloning of the human genomic DNA library, and it was revealed that this gene encoded a protein of 352 amino acid residues, with a calculated molecular weight of 40,600 Da [144]. This protein showed a close association with CCR2, with 75% similarity in both nucleic acid and amino acid sequences [144]. CCR5 was able to bind to several chemokines with different affinities such as macrophage inflammatory protein-1 β (MIP-1 β), MIP-1 α , RANTES (regulated on activation, normal T-cell expressed and secreted), monocyte chemoattractant protein-2 (MCP-2), MCP-3, MCP-4, and MCP-1 [144, 145]. MIP-1 β exhibits the most specific binding to CCR5 while MIP-1 α and RANTES bind to CCR5 as well as several other CC chemokine receptors [119, 146].

CCR5 has been intensively studied because of its role as an HIV-1 co-receptor, which was identified by the discovery of a number of soluble factors released from CD8⁺ T cells that inhibit HIV-1 infection [123, 147, 148]. For example, MCP-2 competes with the HIV-1 gp120 monomer, binding with 80% inhibition efficiency, whereas MIP-1 β inhibits binding by only 30% [145]. In contrast, although MCP-3 exhibits the highest binding affinity for CCR5, suggesting that it might be a natural antagonist for CCR5 and could compete for

HIV-1 gp120 binding, it was shown to be a weak inhibitor of HIV-1 infection because of its inability to downregulate CCR5 expression on the cell surface [145].

The chemokines RANTES, MIP-1 α , and MIP-1 β potently inhibited HIV-1 infection with M-tropic strains of HIV-1 but have shown little effect on infection by T-tropic strains of virus [148]. This study has also shown that cell lines expressing CD4 with one of the CC receptors such as CCR1, CCR2, CCR3, or CCR4 are refractory to infection by M-tropic virus. In contrast, cell lines expressing CD4 with CCR5 are very sensitive to infection by ADA, BaL, or JRFL strains of HIV-1 [35]. Based on these observations, CCR5 has been identified as a co-receptor for primary, M-tropic as well as dual-tropic HIV-1 strains during entry [33–35, 37, 149].

With respect to cellular distribution, CCR5 has been shown to be expressed on the surface of isolated monocytes, macrophages, primary T cells, natural killer cells, and immature dendritic cells. Interaction with a CCR5 ligand is involved in immune cell chemotaxis and cell activation [35, 150, 151]. In the brain compartment, CCR5 has also been found on the surface of perivascular macrophages, microglia cells, and astrocytes [126, 152–154]. Therefore, a virus that has been shown to utilize CCR5 as a co-receptor would be expected to be found in many different tissues within the HIV-1-infected individual and has also been shown to play a role in HIV-1 infection within the brain.

The interaction of HIV-1 with CCR5 during the initial stages of viral entry is conformationally and structurally complex, in part because different viral strains require different and specific domains of CCR5 for viral entry [149]. A study that utilized different dual- and M-tropic strains including JRFL, ADA, SF162, and 89.6 to infect avian cells (QT6) expressing CD4 and different CCR5/CCR2b chimeras suggested that even though CCR5 and CCR2b are both expressed on monocytes and share 76% amino acid sequence similarity, the N-terminus and ECL1 of CCR5 were absolutely required for the entry of both M-tropic and dual-tropic viruses, but M-tropic viruses were the most sensitive to changes in these regions of CCR5 [149]. Truncations of the terminus of CCR5 have shown that the first 2 to 5 amino acid residues of the N-terminus of CCR5 were necessary for the fusion step of M-tropic virus, whereas the dual-tropic virus (89.6) exhibited a strong dependence on the first 6 to 9 residues. These results have suggested that different viruses may utilize the same co-receptor to facilitate viral entry but the specific molecular interactions involved during co-receptor engagement may differ between viruses. The ability of different viral quasispecies to utilize specific co-receptors to enhance the viral entry process may guide changes in viral entry fitness of a given viral variant and hence the evolution of the viral swarm as disease progresses [149].

The role of CCR5 in HIV-1 infection became even more interesting when it was discovered that individuals who were homozygous for $\Delta 32$ -CCR5 were highly resistant to HIV-1 infection [128, 155]. Approximately 1% to 2% of the white population carries the homozygous $\Delta 32$ -CCR5 genetic alteration, and approximately 10% are heterozygous [156, 157]. The homozygous $\Delta 32$ -CCR5 mutation consists of a natural 32-bp deletion in the CCR5 gene in the region corresponding to the second ECL, resulting in the loss of functional CCR5 expression at the cell surface. However, although this allelic variant

resulted in a nonpermissive phenotype with respect to R5 viruses, the mutant allele has not yet been shown to be associated with any detectable phenotype [158]. An individual carrying a heterozygous *CCR5* mutation has a lower level of *CCR5* expression, which has been correlated with a lower level of M-tropic viral replication *in vitro* [158]. Even though it is believed that individuals homozygous for the *CCR5* mutation are highly resistant to HIV-1 infection, in one such individual viral isolates were shown to be T-cell line tropic, phenotypically SI, exclusively CXCR4-using, and able to replicate in primary T cells that are not expressing *CCR5*, approximately 4 years following primary infection [159].

The mechanism of HIV-1 resistance in individuals homozygous for the $\Delta 32$ -*CCR5* mutation is still unclear but may result from a loss in *CCR5* expression or active downregulation of CXCR4 expression by the $\Delta 32$ -*CCR5* mutated protein [160]. $CD4^+$ T cells isolated from homozygous $\Delta 32$ -*CCR5* individuals exhibited a low level of CXCR4 and wild-type *CCR5* along with the endogenously expressed $\Delta 32$ -*CCR5* mutant protein and were less susceptible to CXCR4- and *CCR5*-using, including dual-tropic viruses, as compared with individuals who lacked this mutation [160]. Subsequent experiments have shown that $CD4^+$ T cells expressing the lentivirus-encoded $\Delta 32$ -*CCR5* gene exhibited a high degree of resistance to infection by primary, but not laboratory-adapted, CXCR4-using viral isolates [161]. Furthermore, cell lines expressing a lentivirus-encoded $\Delta 32$ -*CCR5* mutant protein have reduced levels of HIV-1 env-mediated fusion [161]. It has been suggested that $\Delta 32$ -*CCR5* mutated protein altered the stoichiometry of the membrane-associated *CCR5* and consequently perturbed the efficiency of HIV-1 entry, thereby serving as an HIV-1 inhibitory factor. This hypothesis may provide a new avenue for drug development in preventing infection and/or delaying disease progression [128, 161].

Although the absence of productive HIV-1 replication due to defective viral entry *via* interaction with *CCR5* seemed to inhibit HIV-1 infection, thereby slowing disease progression, the detrimental aspects of defective *CCR5* have likely not been fully appreciated, especially in the CNS [151]. In the CNS, chemokines and their receptors are involved in brain development and homeostasis [151, 162]. Several studies have shown that many chemokines and chemokine receptors are expressed in the CNS, including CXCR2, CXCR3, CXCR4, CCR3, and *CCR5* [154, 162–166]. *CCR5* has been studied extensively because of its importance in HIV-1 infection, transmission, disease progression, and HIV-associated neuropathology [154].

Within the brain, studies have shown that *CCR5* is constitutively expressed on microglial cells, astrocytes, and neurons in both normal and inflamed brain tissue [153, 154, 165]. The CNS likely serves as an important viral reservoir; HIV-1 has been detected in brain tissue at the time of autopsy in the absence or presence of neuropathology, with the *CCR5*-using viruses most commonly encountered [154, 167–169]. HIV-1 is thought to first enter the brain early during the course of infection within infected macrophages, which travel to the brain. Moreover, the expression of *CCR5* on brain microvascular endothelial cells might also facilitate CNS entry [170]. Subsequently, HIV-1 has been shown to infect other cells in the brain, most commonly microglia cells, which express both *CCR3* and *CCR5* [165, 171, 172]. The upregulation of *CCR5* expression may also contribute to the spread of HIV-1 within the brain during the course of viral disease [151, 173, 174].

In summary, both CXCR4 and CCR5 have been shown to be involved in the host cell cycle and immune response activation processes; however, they also play crucial roles in HIV-1 infection. CXCR4 has been shown to be a preferred co-receptor utilized primarily by T-tropic viruses. The N-terminus, ECL, and transmembrane of CXCR4 established the vital role in binding to viral gp120, because impairment of these regions has been shown to lead to a significant reduction in HIV-1 infectivity. CCR5 is used predominantly by M-tropic and dual-tropic viruses. The N-terminus and ECL-1 have been shown to be important for gp120 binding; however, slightly different molecular interactions between these regions and gp120 were observed between M-tropic and dual-tropic viruses. CCR5 has become exceedingly important as it is likely involved during HIV-1 transmission and in the genesis of HIV-1-associated neurologic disease.

HIV-1 TROPISM AND PATHOGENESIS

Viral tropism can be defined by the ability of different viral strains or isolates to infect different cell types or tissues and to induce syncytia formation and/or acute or chronic infectious virus production as a result of infection. T-tropic viruses are viruses that can infect T-cell lines and induce syncytia formation in the MT2 cell line, a human T-cell leukemia type 1-infected cell line isolated from cord blood lymphocytes from patients with adult T-cell leukemia [175]. Monocytotropic (M-tropic) viruses have been shown to be able to infect monocytic cells, are unable to infect most T-cell lines, and do not form syncytia [175]. Most, if not all HIV-1 primary isolates are able to infect activated primary CD4⁺ T cells with varying degrees of efficiency; however, the ability to infect T-cell lines and monocytic cells depends on the type of viral strain [175].

The ability to use one and/or more host cell surface molecules as co-receptors for entry is intrinsically linked to the viral phenotypic properties of the virus. The co-receptor families that have been identified include the CC chemokine receptor, the CXC chemokine receptor, and chemokine receptor-like orphan molecules, as described above. HIV-1 preferentially uses either CXCR4 or CCR5 or both as a co-receptor(s) [33–35, 176–178]. In general, the phenotypic nature of viral tropism has been categorized by the use of a specific co-receptor(s) to facilitate viral entry; the CCR5-utilizing (R5) virus designation has been used for a virus that can use the CCR5 co-receptor for viral entry whereas the CXCR4-utilizing (X4) virus designation has been used mainly for viruses that use the CXCR4 co-receptor for viral entry. The dual-tropic (X4R5) virus designation is used for viruses that can utilize either co-receptor molecule during the viral entry process [179, 180].

The tropism for cells of T-cell origin and cells of the monocyte-macrophage lineage has great bearing on the overall nature of HIV-1 pathogenesis and the course of disease (Fig. 3). Although monocytes themselves do not support productive HIV-1 infection *in vitro*, it was demonstrated that HIV-1 could be isolated from mature monocytes, especially from cells expressing CD14 and CD16 [181, 182]. Subsets of cell populations derived from monocytes, such as macrophages and dendritic cells, have been shown to support productive HIV-1 infection [183, 184] and possess a number of biological properties that may impact HIV-1 infection, pathogenesis, and disease including the following: (1) macrophages have a long life span [184]; (2) macrophages have the ability to suppress cytopathic effect resulting from

viral infection and thus prevent cell death [181, 185]; (3) HIV-1 particles can reside in intracellular vesicles or virus-containing compartments (VCCs) within macrophages and dendritic cells, which benefits the virus with respect to evasion of the immune response [186, 187]; (4) cells of the monocyte-macrophage lineage can act as a vehicle for HIV-1 to enter and seed infectious virus within the CNS and thereby infect other cells within the brain [182, 188], (5) cells of the monocyte-macrophage lineage may serve as a cellular reservoir for HIV-1 in the bone marrow, peripheral blood, and tissues during long-term therapy, and (6) cells of the monocytic lineage, specially dendritic cells, may also be key players in facilitating mucosal HIV-1 transmission. T-cell populations also represent a major target for HIV-1 infection and actively produce HIV-1 particles because of the high density of CD4 molecules and the expression of both co-receptors, CXCR4 and CCR5 to different levels depending on the T-cell subset [189–192]. Activated proliferating CD4⁺ T cells are highly susceptible to infection and actively produce infectious virus whereas resting CD4⁺ T cells restrict HIV-1 replication due to the blockage of reverse transcription; however, the presence of several chemokine coreceptor molecules including CCR5 have been shown to facilitate the infection and lead to the establishment of a latent viral reservoir [192–195]. Rather than direct infection, therefore, the resting T cell reservoir is formed by the reversion of activated cells to a resting state [196]. HIV-1 infection causes the death of activated CD4⁺ T cells *via* a caspase-3-mediated apoptosis pathway; in addition, a caspase-1-mediated pyroptosis pathway has been recently proposed to correspond to quiescent CD4⁺ T-cell death [197–199]. The theory of cell turnover during HIV-1 disease progression has also been proposed, according to which, as the loss of CD4⁺ T cells occurs, the numbers of naïve and memory T cells are relatively increased [190]. Especially in HIV-1-infected patients who have undergone successful ART, a normal CD4⁺ T-cell count has been observed because of the increase in the number of naïve and memory CD4⁺ T cells, but not of HIV-1-specific CD4⁺ T cells [191]. Both naïve and memory CD4⁺ T cells exhibit a dominant role during the late stages of disease [200]. Naïve CD4⁺ T cells divide slowly, while memory CD4⁺ T cells maintain the latent reservoir population by undergoing homeostatic cell division; therefore, these cells seem to be responsible for producing infectious viral particles [190, 201].

During HIV-1 transmission, it has been demonstrated that T lymphocytes are the target cells for infection because a transmitted virus preferentially uses CCR5 as a co-receptor, requires a high level of CD4, and is able to replicate efficiently in primary CD4⁺ T cells, suggesting an R5-utilizing T-tropic virus phenotype [202–207]. Subsequently, these transmitted/founder viruses adapt to be able to infect cells of the monocyte-macrophage lineage in order to infect a wider target cell population and one that expresses higher levels of CCR5 than T cells [202]. R5 viruses have gained more importance in HIV-1 pathogenesis because they are able to infect both T cells and cells of monocyte-macrophage lineage. Cells of the monocyte-macrophage lineage serve as reservoirs for chronic infection and have the ability to travel across the blood–brain barrier, contributing to HIV-1-associated neuropathogenesis [188, 207, 208]. It has been reported that the R5 viruses isolated from blood, semen, and different tissue samples such as brain, lung, and lymph node of patients throughout the course of disease exhibited different levels of M-tropic capacity due to the different levels of expression of both CD4 and CCR5 molecules on specific target cell populations [206, 209]. Recent studies utilizing Affinofile cells, which are engineered to express inducible and

titratable levels of CD4 and CCR5 but constitutively express CXCR4, have shown that the M-tropic virus can infect cells expressing low levels of CD4 and high levels of CCR5, whereas the T-tropic virus exhibited less ability to infect these cells [189]. Furthermore, the R5 M-tropic and paired R5 T-tropic viruses exhibited no difference in ability to infect cells expressing low levels of CCR5 and high levels of CD4, suggesting the special ability of R5 virus in utilizing low levels of CCR5 for their target cell infections [189].

The relationship between co-receptor usage and replication capacity in trafficking leukocytes, perivascular macrophages, and microglia, with respect to primary viruses isolated from brain and CSF, have also been investigated to distinguish between CCR5 utilization and M-tropism from the perspective of neurotropism. Similarly, these studies have demonstrated that HIV-1 entry is restricted by a mechanism independent of co-receptor utilization [210]. The highly M-tropic HIV-1 variants isolated from brain tissue have been associated with neurovirulence; however, this is in contrast to virus derived from blood, semen, and lymph node tissue, which require high levels of CD4 for infection [209]. Perivascular macrophages and microglial cells expressing both CD4 and CCR5 are infected and represent a major reservoir of HIV-1 within the brain [188]. The R5 isolates from brain tissue exhibited compartmentalization, were highly M-tropic, and had the ability to induce syncytia formation in primary macrophage cultures, and these characteristics are different from those of R5 viruses isolated from semen, blood, lymph nodes, and spleen [206, 209, 211]. Other cell types within the brain, such as astrocytes and oligodendrocytes, also exhibit HIV-1 susceptibility but do not support virus production [188]. In order to further study neuropathogenesis and the interrelatedness of co-receptor utilization and cellular tropism, the HIV Brain Sequence Database has compiled and annotated HIV envelope sequences sampled from the brain, or from patients from whom the database already contains a brain sequence [212].

During late-stage infection, which involves T-cell loss, the emerging of X4 virus occurs possibly due to the viral adaptation to be able to infect naïve CD4⁺ T cells, which express high levels of CXCR4 [213–215]. In addition, R5 T-tropic virus was found in late-stage disease in lymph nodes and was able to infect memory CD4⁺ T cells and actively produced virus [190, 206, 207], suggesting that during the rapid loss of T cells and progression to AIDS, both the X4 and R5 viruses play a role, but in different populations of T cells.

In conclusion, it would appear that HIV-1 adapts to different cellular compartments during the course of disease with the process of adaptation involving the utilization of different levels of CD4 as well as the differential utilization of co-receptors present on the plasma membrane of different cell populations. In other words, the differential infection profiles are based on the expression levels of cellular CD4, CXCR4, and CCR5 on the target cell populations. T-cell populations play a major role during disease transmission and progression with respect to R5 and X4 viruses, respectively. During clinical latency, the R5 virus predominantly plays a role by virtue of its ability to reside within the MDMs and to infect resting memory CD4⁺ T cells, which contribute to the periodic production of virus in low levels even during ART. Although effective ART has greatly increased life expectancy, HIV-1 reservoirs within the monocyte-macrophage lineage greatly contribute to long-term survival of the virus due to their low turnover [216], resistance to apoptosis [217], and

continuous low-level infection as indicated by continued viral evolution during the course of therapy [218–220]. Among patients on ART, the CD16⁺ monocyte is preferentially infected by CCR5-utilizing strains in addition to being more permissive to replication [221]. Throughout acute and chronic infection, macrophages have demonstrated the ability to evade the immune system through multiple mechanisms, contributing to the inability to clear HIV-1 infection [222]. Cells of the monocyte-macrophage lineage, particularly perivascular macrophages and latently infected monocytes, are thought to be the most important cells with respect to dissemination of the virus throughout the body, including the brain, due to their migratory nature [223, 224].

IN VITRO CO-RECEPTOR USAGE PREDICTION

These viral tropism patterns have been observed in different stages of disease; co-receptor prediction becomes critically important in a clinical context with respect to designing ART regimens. Appropriate therapeutic administration prevents the development of drug resistance and decreased efficacy. Initially, viral tropism was determined by the phenotypic tropism testing (PTT) assay, a method used for verification of the syncytia formation ability of virus in MT2 cells. This test is not particularly well suited for routine clinical application because of its inherent subjectivity and user-defined positive and negative endpoints [225]. The PTT assay was developed using pseudotype virus expressing patient-derived viral envelopes to infect indicator cells that express CD4 and either CXCR4 or CCR5. This test has proved reproducible with respect to determining viral envelope co-receptor usage [225–227]. The original Trofile assay (OTA) and the enhanced-sensitivity Trofile assay (ESTA) are two commercially available PTT assays (Monogram Biosciences, San Francisco, CA) that have allowed for widespread clinical use with respect to defining co-receptor usage by patient-derived HIV-1 envelopes. The ESTA has a higher sensitivity than OTA with regard to detecting X4 virus; however, both commercial PTT assays require a minimum of 1,000 copies/mL of viral RNA to be reproducibly reliable and both are time-consuming, expensive, and technically demanding [225, 228].

Genotype tropism testing (GTT) is an alternative approach that has been used in the prediction of viral tropism and has utilized the env-V3 amino acid sequence of HIV-1 gp120 in conjunction with a number of different computational tools [225, 229]. GTT has several advantages over PTT; it is less expensive, faster, and less technically demanding, and the results obtained have been relatively reliable [225, 230]. Hence, GTT has been more suitable for routine clinical practice [228, 229, 231]. GTT is, however, less sensitive with respect to the detection of rare species of X4 virus, dual-tropic virus, and many different types of clinical isolates [228, 232–234]. GTT initially relied on amplification of the HIV-1 *env* region from plasma viral RNA, a process requiring at least 500 copies/mL to be successful [225]. While the sensitivity of X4 virus detection has also depended on plasma viral RNA levels, an alternative and comparably reliable method for X4 virus detection in patients with low viral load (less than 500 copies/mL) is based on amplification of HIV-1 proviral DNA [225, 235]. Retrieval of the HIV-1 *env* sequence has relied not only on gene amplification but also on the accuracy of the sequencing process utilized. Standard Sanger sequencing has been limited with regard to detecting minority populations of virus, defined as a viral quasispecies population comprising less than 10% of the total population [225, 236]. To this

end, the next-generation sequencing (NGS) approach to DNA sequencing has greatly increased the probability of identifying rare viral populations [228, 237]. Therefore, GTT used in conjunction with NGS has been a promising tool in co-receptor usage prediction because NGS can overcome the detection limitations of viral genome identification, although it has remained an expensive and complex analysis that is not currently widely available [228, 235, 238].

IN SILICO CO-RECEPTOR USAGE PREDICTION

Different regions of the HIV-1 envelope have been studied for determination of co-receptor usage; however, results using sequences from the V3 region of *env* have proven to be the most useful and reproducible [73, 179, 239–242]. Because of the advantages of GTT over PTT, several bioinformatic approaches have been developed to predict co-receptor usage from the V3 sequence. GTT methods can be further categorized as either sequence- or structure-based prediction methods. Sequence-based methods use only information in the one-dimensional amino acid sequence of the V3 while structural methods attempt to model the three-dimensional structure of the V3 loop. Sequence-based methods are typically faster than structure-based methods but encode less information than a three-dimensional analysis.

The earliest genotypic methods relied on amino acids at 11 and 25 in the V3 region; this method is commonly known as the 11/25 rule [73, 243]. These amino acids were selected because research linked them to co-receptor binding during viral entry [73, 231, 244]. According to the 11/25 rule, an X4 virus has a positively charged amino acid at position 11 and/or 25 of the V3 sequence whereas all other amino acids indicated an R5 virus [71]. Early studies showed that amino acids at position 11 in V3 of NSI isolates were likely to be uncharged, such as serine (S) or glycine (G), whereas either uncharged amino acids, such as alanine (A) or glutamine (Q), or negatively charged amino acids, such as glutamic acid (E) or aspartic acid (D), have been frequently found at position 25 [73]. Conversely, in the V3 sequence of SI isolates, positively charged amino acids, such as lysine (K), histidine (H), or arginine (R), have been found at either position 11 or 25 [73]. These charged amino acid substitution strategies have also been shown to be present between M-tropic and non-M-tropic isolates but to a lesser degree [73].

The 11/25 rule was further refined into a “net charge rule” based on a calculation of the total charge of amino acid residues within the V3 loop [73, 74]. The CHARGEPRO program (IntelliGenetics, Hilton Head Isle, SC) implements this algorithm by calculating the V3 loop charge at physiological pH. The total relative V3 net charge of NSI viral isolates is lower than that of SI viral isolates (3.6 ± 0.6 versus 5.5 ± 1.3 , respectively) [73]. Comparing total V3 net charge of M-tropic and non-M-tropic isolates has given similar results (3.6 ± 0.7 versus 4.9 ± 1.2 , respectively) [73]. The net charge rule was further refined in 2000 using the number of positively charged amino acids, the number of negatively charged amino acids, the net charge of V3, and the isoleucine residue at position 292 [74]. This equation resulted in 100% accuracy in co-receptor usage prediction as determined with the given data set of 43 subtype B HIV-1 isolates with known co-receptor use as determined by functional assays [74]. In addition, evaluation of cellular tropism utilizing the total net charge of the V3 loop has demonstrated that NSI variants exhibited values less than +4, whereas SI variants

exhibited values of +5 or greater [74]. The net charge approach yields approximately 91% accuracy with respect to phenotype prediction [74]. The theory of total relative net charge of the V3 peptide may be explained by the fact that charge directly influences the molecular structure or binding capability of viral gp120 [73]. CXCR4 has been shown to be more negatively charged on the surface than CCR5 [245]; therefore, CXCR4-utilizing virus might also have a more positively charged V3 than CCR5-utilizing virus. The net charge rule has exhibited high specificity; however, the sensitivity in defining X4 virus has remained low [74, 246].

The simplicity of the 11/25 and charge rules has increased their use in defining co-receptor usage [73, 243]. Amino acid charge has appeared to relate to the secondary structure of the V3 loop, which has resulted in uniquely structured V3 loops of M-tropic and non-M-tropic isolates [73]. The 11/25 rule predicts X4 utilization with approximately 60% sensitivity and 92.5% specificity [231]. In combination with other computer-based technologies, the prediction of sensitivity and specificity values can be increased [231]. This advance has resulted in the development of new algorithms such as geno2pheno and WebPSSM (described below in more detail) based on a combination of the 11/25 and the V3 net charge rules. This algorithm has considered amino acids at positions 11 or 25 (containing K or R), as well as the total net charge of V3 (+5 or +6), and has yielded a concordance between genotype and phenotype as high as 91% from plasma virus samples (using samples from 103 infected patients, 85% were shown to be subtype B, and the remaining were A, F, G, J, CRF-01, CRF-02, and CRF-06 subtypes) [247]. Although co-receptor prediction based on the 11/25 and charge rule gives more than 80% overall accuracy, this leaves a false prediction rate as high as 20%, leaving ample room for improvement. The next section will describe the three Web-based bioinformatics tools that have been developed and are currently available to enhance predictive capabilities (Table 1). These tools are limited to the utilization of linear V3 sequences; nevertheless, they employ different methods of analysis as well as different data sets [231].

Wet-Cat

Wet-cat (<http://genomiac2.ucsd.edu:8080/wet-cat/v3.html>), the oldest Web-based bioinformatic tool for co-receptor usage prediction, was developed at the University of California at San Diego in 2003 [248]. Wet-cat has used CLUSTALW (University College Dublin, Ireland) to generate an alignment between a query sequence and V3 sequences from the Los Alamos National Laboratory (LANL) HIV sequence database [248]. Classifiers including Quinlan C4.5, partial decision trees constructed with C4.5 (PART), and support vector machines (SVM) have been trained to predict binding to the CXCR4 co-receptor [248]. Among the machine learning techniques, SVM was the most effective when performing 100 rounds of 10-fold cross-validation with the complete V3 sequence [248].

Wet-cat has compared the ability of different machine learning techniques to a simple charge rule and has shown that the entire sequence within V3, rather than just amino acid positions 11 and 25, contributes to co-receptor determination, and these machine learning techniques exhibit nearly 90% accuracy in prediction [248]. Exclusion of amino acid positions 11 and 25 decreased the accuracy by approximately 5%, with position 11

contributing the majority of the lost information [248]. Overall, these results suggest that all amino acids within the V3 region contribute to co-receptor usage selection. In addition, further studies involving HIV-1 envelope mutagenesis have indicated that amino acid residues outside the V3 loop region could influence the structure of gp120 induced by CD4 binding, which is critical for co-receptor binding and viral tropism [248, 249].

Similar to other *in silico* prediction methods, wet-cat is insensitive when predicting a group of dual-tropic viruses [248]. Because of the nature of binary classification systems, most tools have been trained to predict binding to the CXCR4 co-receptor and as such classify dual-tropic viruses as CXCR4 binding as opposed to CCR5 binding. In summary, wet-cat has provided another method of co-receptor usage prediction by which a user can select known trained classifiers in the prediction process; however, wet-cat has relied solely on reference sequences from the LANL database and has centered on the sole use of the V3 sequences within the 34 to 36 amino acid range with respect to length, and therefore sequences outside this region have not been included in comparative analyses using wet-cat [248]. Because this approach relies heavily on LANL, care needs to be taken in the future when more sequences are obtained from HIV-1-infected patient populations that LANL or an equivalent database is continuously updated with respect to the breadth of sequences and knowledge in the field.

PSSM

Another bioinformatics tool that has been applied to the co-receptor prediction problem is the position-specific scoring matrix (PSSM), which has been implemented in the WebPSSM tool (<http://indra.mullins.microbiol.washington.edu/webpssm>). This tool focuses on only amino acid sequences from the V3 loop [235, 250]. Unlike other *in silico* co-receptor usage predictions, the PSSM algorithmic method offers not only co-receptor prediction but also a measurement of viral phenotype transition *via* a defined scoring system [251]. PSSM evaluates a V3 sequence by aligning it against subtype B sequences with known phenotypes derived from functional experimentation in the training data set, which were retrieved from [251–253]. This sequence data set was composed of both SI/NSI and X4/R5 data sets [252]. A V3 sequence comparison calculates a score using the Needleman-Wünsch algorithm and amino acid matrix, which distinguishes between sequences derived from X4 and R5 viruses [235, 251]. The scoring method has been shown to effectively evaluate the similarity of an input sequence to known X4 V3 sequences and generates a matrix of ratio scores of each amino acid position; hence, a higher score indicates the likelihood that the V3 sequence has been derived from an X4 virus [254]. The score distribution has been shown to predict R5 viruses as those with V3 sequences having score values below -6.96, whereas V3 sequences with score values higher than -2.88 were predicted to be derived from CXCR4-using (X4) viruses [251]. The overlapping scores between -6.96 and -2.88 have been shown to be V3 sequences potentially derived from dual-tropic viruses, which have been shown to be an evolutionary intermediate between X4 and R5 viruses [251] or a V3 sequence derived from atypical X4 or R5 viruses. A majority of V3 sequences in the PSSM training data set are 35 amino acids long; any other length V3 sequence requires a multiple alignment and co-receptor prediction for each alignment [251].

PSSM has two different matrices; one has been designated X4/R5, which aligns the query to a set of 213 HIV-1 subtype B V3 sequences with known co-receptor phenotypes [246, 251, 252]. A secondary matrix designated SI/NSI can be used to determine a score from a set of 257 viruses with known SI phenotypes identified using the MT2 cell line [251, 252]. Furthermore, PSSM currently offers a co-receptor prediction for subtype C viruses (PSSM-C) but only for the SI/NSI matrix [255].

HIV-1 subtype C is responsible for approximately half of all global infections, although X4-utilizing viruses have been shown to be rare within this population [255–258]. Examination of HIV-1 subtype C V3 sequences has shown that, unlike subtype B, positively charged amino acid substitutions at positions 11 and 25 were not correlated with the SI/NSI phenotype [255, 258]. Consequently, four different methods were compared in subtype C phenotype prediction, including the 11/25 rule; Briggs (which has been referred to as a multiple regression method or total V3 net charge); SVMs [259]; and B-PSSM (SI/NSI). It is not surprising that these methods yielded poor performance on subtype C sequences, especially the NSI phenotype, because they were originally trained based on subtype B sequences [255]. Nonetheless, the PSSM algorithm was the most effective of the four subtype B predictive tools with respect to predicting both subtype C SI and NSI variants [255]. PSSM-C was subsequently developed based on 279 subtype C V3 sequences, each containing 35 amino acid residues with known syncytium capability.

Among the HIV-1 subtype C V3 sequences used in training PSSM-C, 228 NSI sequences (from a total of 200 individuals) and 51 SI sequences (from a total of 20 individuals) were retrieved from the published literature, the Entrez nucleotide database, the LANL HIV sequence database, and other unpublished data [255]. The SI/NSI phenotypes of V3 sequences within this subtype C training data set were mostly identified using phenotyping methods from bulk viral preparations, not from clonally derived viruses; consequently, the results must include this caveat when drawing or extending conclusions because of the heterogeneity of each of these viral populations [255]. However, this approach is supported by previous studies utilizing a subtype B training data set that demonstrated little difference between PSSM prediction utilizing bulk viral preparations versus cloned isolates, likely because of the large size of the data set employed in this analysis [255]. Subsequently, the performance of the PSSM-B and PSSM-C algorithms with respect to determining phenotypic predictions of subtype C V3 sequences was examined. Based on a receiver operating characteristics analysis, these studies revealed that both PSSM algorithms exhibited insignificant differences in predicting subtype C phenotype, indicating that PSSM-B could be useful in this regard. However, with optimization of the PSSM cutoff, the sensitivity of the test was improved with limited decrease in specificity [255]. To validate the new optimized PSSM-C training algorithm, a different set of unique subtype C sequences was tested and compared with PSSM-B with a modified subtype C cutoff. PSSM-C yielded 83% specificity (39/47) and 83% sensitivity (20/24) sensitivity while PSSM-B with modified subtype C cutoff exhibited 70.7% specificity (37/47) and 54.1% sensitivity (13/24) [255]. These results clearly indicated that PSSM-C performed better with respect to HIV-1 subtype C phenotype prediction, although there remains room for improvement in this regard.

Utilizing tools like the PSSM algorithm, a study of co-receptor usage determination in HIV-1-infected patients receiving Maraviroc through an expanded access program in Europe compared several V3 genotypic methodologies with the Trofile phenotypic assay [260]. Sensitivity in X4 virus detection was low, ranging from 31% to 75% [260]. Based on these results, two new PSSM algorithms, PSSM_{X4R5-8} and PSSM_{SINSI-6.4}, were developed in order to enhance the sensitivity of X4 variant detection [260]. The new PSSM algorithms have been optimized, resulting in the development of new threshold levels, -8 and -6.4 for X4/R5 and SI/NSI determination, respectively [260]. Consequently, utilizing these thresholds, a sample with a PSSM_{X4R5} score higher than -8 is considered to be X4, and a sample with a PSSM_{SINSI} score above -6.4 is considered to be an SI virus. In addition, both algorithms have a sensitivity level as high as 80% with respect to detecting X4 virus [260]. PSSM_{X4R5-8} and PSSM_{SINSI-6.4} algorithms were also evaluated in a separate set of 148 HIV-1-infected plasma samples from the ALLEGRO trial (a multicenter study in Spain). These studies demonstrated an increased sensitivity (93%) with respect to the detection of X4 virus in both algorithms [260]. However, specificity of both sample sets was low, along with positive predictive value (approximately 50%, indicating that half of the sample predicted as X4 could be R5 by phenotypic assay) [260]. Therefore, there is likely a trade-off between sensitivity and specificity with respect to detecting X4 variants between HIV-1-infected patients on ART, suggesting that the predicted X4 samples should be reconfirmed using a phenotypic assay [260].

Geno2pheno

Geno2pheno is another Web-based tool that has been widely used for co-receptor prediction. Geno2pheno was established by a group of scientists in Germany in 2002 and currently provides several Web-based medical services such as HIV-1 gag, pol, integrase, and drug resistance prediction (<http://www.geno2pheno.org>). The Geno2pheno algorithm is based on analysis of phenotype-genotype corrections using SVM and decision trees to implement the range of different cut-offs used in data interpretation [253, 261, 262]. SVMs have been considered the best-trained phenotype classifiers, outperforming decision trees and neural networks, primarily because of their consistency and precision with respect to interpretation of data [248, 263]. Geno2pheno converts genotype into phenotype by generating multiple alignments against 1,100 known reference sequences of V3-containing representatives from all types of co-receptors, which were retrieved from the Los Alamos database as well as from the literature [253, 263]. Included with the 1,100 reference sequences were 769 R5 sequences, 210 X4 sequences, and 131 dual-tropic sequences, from a total of 332 patients [232].

Geno2pheno has gained widespread usage because it has several analytic advantages. First, geno2pheno exhibits 86.5% and 79.7% concordance with the phenotypic assays for co-receptor use, Trofile and Tropism Recombinant Test, respectively [227]. Second, it is easy to use because the V3 sequence can be uploaded as either plain text or in FASTA format. A group of sequences can be submitted; the program will analyze and generate data output in table format, which can be downloaded in comma-separated-value format that is machine readable for downstream analysis. The level of conservation relative to CXCR4 usage prediction can also be adjusted by the user. This strategy has also been useful in the

prediction of CCR5-utilizing virus in clinical practice to determine when a CCR5-blocking therapeutic agent can be effectively administered.

Another advantage of geno2pheno is that the input V3 sequence can not only be analyzed based on standard clonal population data but can also be applied to a clinical data set (275). The standard clonal population data have been collected by cloning of individual viral isolates obtained from 1,100 samples derived from 332 patients (250, 251). Analysis of individual clinical samples based on population-based sequencing results showed that the specificity values are close to those observed in the analysis of the clonal sequence data set; however, the sensitivity of the population-based sequence determinations was decreased [232]. Therefore, geno2pheno offers the use of additional clinical parameters in determining co-receptor usage. The clinical analysis model has been based on V3 sequence data that were previously determined by population-based bulk sequencing with co-receptor phenotype determined by the Trofile co-receptor assay, from 952 ART-naïve patient samples including the use of clinical parameters in co-receptor use prediction [264]. Utilization of clinical parameters such as CD4⁺ T-cell count, CD8⁺ T-cell count, and/or viral load in co-receptor usage prediction based on a population-based sequence data set has been used to improve the sensitivity of CXCR4-utilizing virus detection performance, and this combination employs the 11/25 rule [232]. Because clinical samples are comprised of a mixture of quasispecies, population-based sequencing has become useful in identifying the predominant genotype, and the traditional cloning combined with clone sequencing has slowly been replaced by ultradeep sequencing or single-genome sequencing [231, 265, 266].

STRUCTURE-BASED PREDICTION METHODS

A three-dimensional model of the V3 loop inherently encodes more biologically relevant information than a linear sequence. However, only recently has there been enough three-dimensional structural data, with the first structure being published in 2009 [76], to utilize this information for tropism prediction. Currently three popular prediction methods utilize structural information: Sander *et al.* [267], Dybowski *et al.* [268], and geno2pheno-structure [269]. Of these three methods, only geno2pheno-structure has public Web server implementation.

The greatest problem with using a structural prediction method is that predicting patient tropism requires translating an amino acid sequence into an approximate crystal structure. The prediction method by Sander *et al.* (2007) uses the Side Chains with a Rotamer Library (SCWRL) model [270] to rapidly predict the side-chain arrangement of the residues in the V3 model. The pitfall of this method is that it cannot handle insertions, restricting this method to the analysis of only V3 sequences free of insertions or deletions. The pairwise distances between all pairs of donor atoms, acceptor atoms, aliphatic atoms, etc., as classified by [271], were used as feature vectors in an SVM model. On a set of 514 distinct clonal V3 sequences extracted from LANL, this method was compared to the 11/25 rule. Utilizing this data set, the 11/25 rule had a specificity of 0.9463 and sensitivity of 0.6186, while the Sanders *et al.* (2007) prediction method achieved an identical specificity with a sensitivity of 0.7340. Combining the linear sequence information further increased the sensitivity to 0.8041.

An alternate method proposed by Dybowski *et al.* (2010) uses Modeller v9.6 [272] to estimate the three-dimensional structure for unknown clonal sequences. An electrostatic hull, which is the electrostatic potential for every position on a 0.6-nm grid, was calculated using an adaptive Poisson-Boltzmann solver [273]. This massive feature set was reduced by using random forests [274] and combined with a similar random forest generated from hydrophobicity of the linear sequence using a scale developed by Kyte and Doolittle [275]. These two feature sets were used to generate a second-level random forest model trained to distinguish between X4 and R5 V3 sequences. This two-level, state-of-the-art model obtained a sensitivity of 0.81 with a fixed specificity of 0.97 and an area under the concentration curve (AUC) of 0.937. This was accomplished on a set of 1151 R5 and 166 X4 clonal sequences sampled from multiple subtypes. Although this method is able to achieve a stunningly high sensitivity, it is not able to increase the understanding of the binding dynamics of the V3 loop. The random forest method obscures the relationship between the position in the three-dimensional structure and its contribution to the resulting score.

The researchers who developed the geno2pheno algorithm, mentioned above, have begun to integrate structural information into their prediction algorithm [269]. This is the only method that has a Web-based implementation (<http://structure.geno2pheno.org/>). It uses a binning method similar to the method described by Dybowski and coworkers; however, instead of solving the extensive electrostatic equations in the entire hull of the V3, it simply calculates the properties of each amino acid in 56 preselected positions. This significantly simplifies the calculation compared with the models proposed by Sanders *et al.* (2007) and Dybowski *et al.* (2010). On a data set of 1186 subtype B V3 clonal sequences from the LANL database, the geno2pheno algorithm obtained a sensitivity of 0.587 for a fixed specificity of 0.97 and an AUC of 0.847. Further testing of this method on identical data sets to the methods described by Sanders *et al.* (2007) and Dybowski *et al.* (2010) was completed. Geno2pheno outperformed the method described by Sanders *et al.* (2007), with sensitivities of 0.782 and 0.774, respectively, and the method described by Dybowski *et al.* (2010), with sensitivities of 0.838 and 0.810, respectively. When tested on the HAART Observational Medical Evaluation and Research (HOMER) data set [276], which contains sequences derived directly from patient samples along with clinical parameters, the model achieves a sensitivity of 0.463.

PITFALLS OF CO-RECEPTOR USAGE PREDICTION

Viral tropism has been shown to evolve during the course of HIV-1 disease. As previously detailed, in the early stage of HIV-1 subtype B infection, the majority of cloned viral isolates are R5 viruses. The R5 virus also dominates during clinical latency or asymptomatic disease and during the later stages of disease when the T-cell compartment has been irreversibly depleted. The X4 virus has been shown to be the most prevalent virus during the course of symptomatic disease and has been linked to CD4⁺ T-cell depletion [277, 278]. Analysis of ART-naïve patients has shown that the prevalence of X4 virus was less than 10% when CD4⁺ T-cell count was more than 200 cell/ μ L, but the number of X4 viruses increased significantly when CD4⁺ T-cell count dropped below this point, when plasma viral load was significantly high, and these patients were most likely to have an AIDS-defining illness

before starting ART [279]. On the other hand, among ART-experienced individuals, patients who carry only X4 virus have been shown to be rare; however, X4 virus has been shown to co-exist with R5 and dual-tropic viruses at very low levels (2%–7%) in patients exhibiting a range of CD4⁺ T-cell counts, even relatively high counts [280–282]. Dual-tropic and R5 viruses are commonly found in ART-experienced patients, as compared with those who have either had limited treatment or are ART naïve; the emergence of dual-tropic virus may result from therapeutic selective pressures as observed in 27% to 58% of individuals who had received ART with a range of CD4⁺ T-cell counts [231, 280]. The prevalence of X4 and dual-tropic viruses has also been shown to be independent of CD4⁺ T-cell count and CCR5 32 genotype [283]. Therefore, it has become clear that X4 and dual-tropic viruses could arise any time during the course of disease, but the low sensitivity with respect to the detection of both X4 and dual-tropic viruses in clinical samples has remained problematic among the currently available co-receptor usage prediction algorithms.

Trofile and other co-receptor phenotyping assays are relatively reliable, offering robust viral tropism characterization panels with respect to X4-, R5-, and dual-tropic viruses; however, they are expensive, time consuming, and inconclusive in approximately 15% of clinical samples [260, 284]. Application of genotyping technology as a screening tool could be beneficial and could substantially decrease the need for phenotypic assays.

Several currently available *in silico* co-receptor prediction methods have provided comparable results [232, 234, 253, 281, 285, 286]. However, one study was performed with co-receptor predictions based on 83 plasma samples from HIV-1 subtype B-infected individuals, utilizing three different *in silico* co-receptor prediction methods (wet-cat, PSSM, and geno2pheno) compared with a phenotypic assay [285]. The overall results indicated that PSSM gave the best concordance with the phenotypic assay, followed by geno2pheno and wet-cat algorithms [285]. The 11/25 rule is still favored based on its simplicity and more than 90% accuracy in prediction [73, 287, 288]. The V3 net charge has not been used alone quite as widely as the other genotyping algorithms, and its use is also complicated because of the ambiguity of the H (histidine) residue in charge calculation methodology [74, 250, 287]. The V3 net charge has often been used in combination with the 11/25 rule, and this approach has increased the specificity of the assay up to 98% in both subtype B and non-subtype B, but it is rather insensitive [247, 289].

Both GTT and PTT tropism methods have been utilized with respect to clinicians prescribing Maraviroc to HIV-infected patients. Maraviroc has a tropism-specific effect and inhibits only R5 tropic viruses [290]. A recent study used Trofile and geno2pheno to determine tropism in a set of 25 patients and found that GTT methods led to a misclassification rate of only 16.7% [291]. Another study found that in a collection of 145 clinical samples the WebPSSM prediction method had the highest agreement with the Trofile assay, with a 17.4% misclassification rate [292]. Finally, in the *Agence Nationale de Recherche sur le SIDA* (ANRS) resistance group study, a set of 189 patients was tested using both Trofile and the geno2pheno algorithm. Likely because their viral infection was drug resistant, in these patients the geno2pheno algorithm misclassification rate jumped to 60.7% [293]. These studies are likely the first of many examining the classification rate in clinical samples, which present unique challenges relative to clonal sequences.

Although a majority of predictive tools have been constructed based on HIV-1 subtype B, with the exception of PSSM-C, which offers only an SI/NSI matrix, viral tropism prediction in non-subtype B viruses has remained incomplete as compared with subtype B [281, 287, 289]. Considering only HIV-1 subtype B analysis, PSSM exhibited better concordance with respect to genotyping predictions, and applying the threshold score of -8 enhanced the sensitivity of X4 detection [281, 285]. However, the PSSM algorithm is error prone in predicting co-receptor usage from any V3 sequences with more or less than 35 amino acid residues [253]. Geno2pheno has also been shown to exhibit a substantial level of accuracy in tropism prediction, but its accuracy is lower for detecting X4 or dual-tropic viruses and it is significantly less sensitive with respect to determining co-receptor usage of clinical isolates [232, 253]. The use of the geno2pheno algorithm has been ambiguous based on the false-positive rate (FPR) adjustability because each cutoff has conferred a different level of sensitivity with respect to X4 detection [228, 294]. High sensitivity and specificity with respect to the detection of X4 viruses could be demonstrated by applying an FPR more than 2.5% in both subtype B and non-B analyses; however, an FPR cut off at 10% has been recommended to provide better overall sensitivity and specificity [287]. Viral tropism analyses within non-subtype B viruses have shown that PSSM has greater sensitivity and specificity in subtypes A, B, and G, but not in subtype F (Table 2) [287]. The combination of classic 11/25 and V3 net charge rules has also been shown to be a good candidate with respect to predicting the co-receptor usage of non-B viruses with a high degree of specificity (over 90%); nevertheless, sensitivity in detecting the X4 or dual-tropic virus has been more uncertain, depending on the subtype [287, 289]. The combined 11/25 and V3 net charge algorithm was utilized in co-receptor prediction analysis of HIV-1 CRF01-AE and the level of specificity was greater than that of geno2pheno but the sensitivity was much lower [289]. Subsequently, another simple new rule for HIV-1 CRF01-AE co-receptor prediction was established by incorporating a newly identified N-linked glycosylation determinant within the V3 loop into the current combined 11/25 and net charge rule, as it was noticed that the X4 CRF01-AE exhibited the loss of the N-linked glycosylation site at the beginning of the V3 loop [289]. This new algorithm performs better (70% sensitivity and 96% specificity) for X4 prediction in HIV-1 CRF01-AE [289]. Similarly, the original combined algorithm involving the 11/25 and net charge rules has been applied to an analysis of viral tropism from a total of 32 subtype D-infected individuals; it yielded poor specificity (68%) although the sensitivity was much higher (100%) [295]. The new combined rule specific for subtype D methodology was subsequently adapted because of the frequently observed lysine amino acid at position 25 of the R5 V3 loop sequence [295]. The new combined 11/25 and net charge rules for subtype D gave a better performance (92% concordance with 75% sensitivity and 95% specificity) [295]. This new subtype D combined algorithm also showed an improved viral tropism prediction exhibiting 85% concordance with the phenotype assay, with 68% sensitivity and 95% specificity, as compared with geno2pheno and the original combined 11/25 and net charge algorithm, validated by an independent data set of 67 subtype D sequences retrieved from GenBank [296].

In summary, a majority of current bioinformatic co-receptor usage prediction tools have been constructed based on HIV-1 subtype B V3 genotype amplified from viral RNA or proviral DNA with more than 80% accuracy. While structure-based methods may improve

sensitivity, the higher implementation cost and lack of studies on their performance on clinical data sets imply that they cannot be solely used to predict viral tropism. Discordance in co-receptor prediction often occurs because of the heterogeneity of clinical samples, subtypes, and choice of algorithm, suggesting that in routine clinical practice, multiple algorithms should be used in evaluation of co-receptor usage and tropism.

BEYOND THE V3 LOOP WITH REGARD TO GENOTYPING AND VIRAL TROPISM

While many co-receptor algorithms favor the application of V3 sequence in prediction, utilization of other regions within gp160 has also been proposed [179, 286, 297]. Study of the full length of the envelope region has shown that specific nucleotide positions represent different levels of predictive value after utilizing a TAN (tree-augmented naïve Bayes) classifier with a feature selection algorithm [286]. A newly developed TAN classifier has incorporated a total of 26 selected nucleotide positions that have been located within both identified and unidentified regions of gp120 and gp41, demonstrating a total accuracy of 97.8%, specificity of 100%, and sensitivity of 84.6% in co-receptor prediction, validated by *in vitro* co-receptor assays [286]. Compared with the geno2pheno and PSSM algorithms, the TAN classifier has been shown to be more accurate, more specific, and more sensitive in both subtype B and non-subtype B tropism analyses [286]. Based on the fact that HIV-1 subtype B infection represents only 11% of global infections, non-subtype B tropism prediction tools are needed because these infections are increasing in prevalence worldwide [298]. Utilizing a full-length gp160 nucleotide sequence in tropism prediction has been thought to be ideal based on its accuracy, specificity, and sensitivity in all non-subtype B analyses; nevertheless, a disadvantage centering on sequence quality might be encountered. Another proposed strategy with respect to the improvement of co-receptor usage termination has centered on incorporating the V2 genotype into a current algorithm that utilizes V3-specific characteristics [179]. Analysis of the V2 loop sequences (subtype was not indicated) that have been retrieved from the Los Alamos HIV Sequence Database showed that several mutations and position-independent features are different between X4 and R5 viruses [179]. Therefore, a new algorithm has been developed by utilizing different identified features of the V2 loop, a set of V3 clonal sequences, and SVMs, showing that employing both V2 and V3 regions increases the overall prediction performance, tested in both therapy-naïve and therapy-experienced individuals [179]. Additional studies within gp41 demonstrated that mutations within gp41 exhibit a correlation with co-receptor usage [297]. However, incorporating these mutations into the current prediction model, which utilizes V3 sequence alone, did not improve co-receptor usage prediction, validated with two different cohorts that contained more than 2,000 genotype/phenotype clinical samples [297].

CO-SELECTION OF AMINO ACIDS WITH V3

As indicated previously, viral tropism can be influenced by either V3 alone or a combination with other viral genotypes. To examine whether the V3 sequence of X4 and R5 viruses exhibits different genotypic patterns in other regions of the viral genome, mutual information (MI) [299] was utilized to identify four regions that contain co-evolved sequences with the V3 sequence. Applying this method to 2,010 subtype B sequences, and

937 subtype C sequences that have a V3 sequence along with a co-linear gp41, gp120, LTR, and Nef sequences retrieved from the LANL database. These sequences were subsequently divided into two populations (X4 and R5) utilizing the PSSM algorithm [250]. The MI was calculated between all positions within the V3 region and all positions within gp120, gp41, LTR, and Nef, and subsequently evaluated with 10^6 random permutations to estimate the statistical significance [299]. This process was applied for each subtype (B and C) and co-receptor subset separately, and shown in Fig. (4) are positions that pass a Benjamini Hochberg correction at $p < 0.05$ [300].

In general, it was observed that the V3 sequence was highly linked to the splice site between gp41 and gp120, implying that this linkage was not related to the co-receptor phenotype (Fig. 4). The amino acid segment from 240 to 260 within gp41 was linked in the combined subtype B and C analysis, but not in the individual subtype analyses, implying that the region was linked with the B and C genotypes but not the viral tropism. Interestingly, the gp120 region 85 to 100 was linked with the X4 phenotype in all subtypes, indicating a link between the X4 phenotype and the C1 region.

DIFFERENTIAL AMINO ACID DISTRIBUTION BETWEEN X4 AND R5 PHENOTYPES

Using the same subtype B and C sequence populations described above, a Fisher's Exact test was performed to determine positions that exhibited a differential amino acid distribution between X4 and R5 phenotypes. After sequences were separated into X4 and R5 phenotypes using the WebPSSM tool [250], each position was compared back to the ConB (consensus subtype B) reference sequence using the Muscle algorithm [300], and a Fisher's Exact test was used to find regions that had a statistically significantly different proportion of amino acid substitutions between predicted X4 and R5 sequences. Positions that passed a Benjamini Hochberg correction at $p < 0.05$ [300] are shown in Fig. (5) and Table 3. While a majority of studies have demonstrated viral tropism within the viral envelope, for the first time, this study has shown that regions outside the envelope exhibited linkage with co-receptor utilization in both subtypes B and C (Figs. 4, 5 and Table 3).

Cross-Talking within the Viral Genome

Viral adaptation occurs in response to several cellular selective pressures and has been shown to correlate with the changes throughout the entire viral genome [301–303]. Correspondingly, previous studies have shown that changes in V1/V2 as well as C4 loop regions influenced viral tropism both alone as well as in conjunction with changes within V3. A prior study [304] as well as results presented herein (Figs. 4, 5 and Table 3) demonstrated that increased variability at amino acid positions 190 and 440 is associated with the X4 phenotype. Within gp120, outside of V3, amino acid position 440 showed the most dramatic linkage with an inferred phenotype. The linkage between the V3-derived phenotype and position 440 has been characterized in a number of studies [241, 304, 305]. The relationship between V3 and C4 (which contains position 440) has been characterized as both physical and functional [306–308]. These studies suggested that the amino acid substitutions at position 440 may have an impact on gp120 tertiary structure by minimizing

the amount of electrostatic repulsion between V3 and C4 regions, therefore associating with co-receptor binding and antibody recognition [306–308].

Amino acid position 190 located within the V1/V2 stem has been shown to be a position at which higher variability occurs among X4 variants [304]. This position has been described as the first well-conserved amino acid after the region of extensive length polymorphism in V2, and, it is part of a motif that has been predicted to direct N-linked glycosylation of position 188N [304]. The position 190 amino acid substitution identified within the X4 virus was frequently positively charged, thus increasing a total V2 net charge, and possibly interrupting the glycosylation motif [304]. Furthermore, the serine amino acid that has often been observed at position 190 (190S) of the X4 virus was identified in a virus isolated from a patient and adapted to grow *in vitro* in microglial cells [309]. The virus containing 190S exhibited the increased level of fusogenicity of the CCR5-utilizing virus, which could explain the mechanism of viral spread within the brain and the development of HIV-1 associated dementia [304, 309].

Additionally, our analyses (Figs. 4, 5 and Table 3) have shown multiple linkages between V3 and gp41 residues. This is intuitive given the complex these two proteins form during oligomerization within the HIV-1 envelope as well as their interaction and conformational change during HIV receptor and co-receptor engagement, as well as fusion events during viral entry. Indeed, a number of studies have also found that selected amino acid variations within gp41 may associate with co-receptor utilization [240, 297]. One study demonstrated that, especially within the context of X4 virus, there were a number of statistically significant associations, and almost all of these associations were amino acids localized within the extracellular domain and exposed on the surface of the glycoprotein [310]. This study hypothesized that these particular amino acid residues might participate in the stabilization of the gp120-gp41 complex and therefore influence the co-receptor binding step [310]. Another study determined that mutations in the fusion peptide and cytoplasmic tail of gp41 contributed to CXCR4 use by a dual-tropic clone, while a single glycine to valine change at position 515 (G515V) within gp41 fusion peptide of another dual-tropic clone was sufficient to confer CXCR4- to the CCR5-utilizing phenotype [240]. This also held true for subtype C virus in addition to subtype B viruses [311]. In addition to determining gp41 sequence involvement in co-receptor utilization, other studies have determined that three conservative changes in the fusion peptide of gp41 resulted in resistance to the small CCR5 inhibitor Vicriviroc [312] and other mutations within both gp120 as well as gp41 resulted in modulation of the magnitude of resistance to aplaviroc (an CCR5 antagonist) [313].

Our studies presented here have also shown a number of variations within Nef as well as a single variation within the LTR to correlate with the V3 genotype. Nef and Env have a few overlapping functions, including the ability to reduce surface expression of CD4 [314, 315], the ability to alter CD3 signaling [316, 317], as well as the ability to alter apoptosis [318]. This result has suggested that there might be a correlation between Nef sequence and co-receptor utilization, however; one study suggested that while Nef did enhance viral infectivity, it did so independently of co-receptor tropism [319]. More research needs to be pursued, nonetheless, to fully elucidate any role of Nef related to co-receptor utilization. To date, no studies have demonstrated a link between the LTR and the V3 region.

SUMMARY

HIV-1 tropism can be classified in part by co-receptor usage during the viral entry step. X4 and R5 viruses are defined by their use of CXCR4 and CCR5 chemokine co-receptors, respectively. Because current HIV-1 therapeutic protocols have often utilized a CCR5 inhibitor in patients infected with only R5 virus, it has been important to verify viral co-receptor usage in patients prior to treatment [320]. Many bioinformatic tools have been successfully developed to ease the complications of viral tropism phenotypic assays, and these tools have been shown to be reliable with results comparable to *in vitro* phenotype assays. The original basic idea of using an algorithm was to utilize the different V3 genotypic differences between X4 and R5 viruses with an array of computational tools to identify specific viral phenotypes with respect to co-receptor utilization. Because a majority of currently available *in silico* co-receptor usage predictive tools are constructed based on HIV-1 subtype B V3 genotypes, applying these to non-subtype B viruses will likely be less effective in predicting co-receptor utilization. Individual algorithms display different levels of specificity and sensitivity with respect to co-receptor usage prediction in different HIV-1 subtypes, and it is essential to apply multiple algorithms to each analysis. Because of the complications of viral quasispecies as well as the imperfection of current bioinformatic tools, the co-receptor genotypic assays remain less capable of detecting dual-tropic viruses. Thus, many more algorithms are being developed, by incorporating, for instance, other characteristics within viral gp120 and gp41 regions, including improvements with respect to sequence generation.

While the accuracy of existing co-receptor *in silico* programs must continue to be assessed, what has been missing is the utilization of genetic variation in other areas of the HIV genome that may relate to differential cell utilization, immune escape, and/or HIV-1 disease severity. The analyses presented for the first time herein involve novel linkages between the HIV-1 V3 sequences and other parts of the viral genome. Understanding these areas of the genome with respect to envelope utilization through examination of co-linear sequences, or other yet unidentified mechanisms, will be important for design of future co-receptor usage predictive technologies and the development of therapeutics directed at entry mechanisms.

Acknowledgments

These studies were funded in part by the Public Health Service, National Institutes of Health, through grants from the National Institute of Neurological Disorders and Stroke, NS32092 and NS46263, the National Institute of Drug Abuse, DA19807 (Dr. Brian Wigdahl, Principal Investigator), National Institute of Mental Health Comprehensive Neuro AIDS Core Center (CNAC), P30 MH-092177 (Kamel Khalili, PI; Brian Wigdahl, PI of the Drexel subcontract), and under the Ruth L. Kirschstein National Research Service Award 5T32MH079785 (Jay Rappaport, PI, Brian Wigdahl, PI of the Drexel subcontract). The contents of the paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Dr. Michael Nonnemacher was also supported by faculty development funds provided by the Department of Microbiology and Immunology and the Institute for Molecular Medicine and Infectious Disease.

References

1. Gallo RC, Montagnier L. The discovery of HIV as the cause of AIDS. *N Engl J Med*. 2003; 349:2283–5. [PubMed: 14668451]
2. Clements JE, Zink MC. Molecular biology and pathogenesis of animal lentivirus infections. *Clin Microbiol Rev*. 1996; 9:100–17. [PubMed: 8665473]

3. Anderson E, Zink W, Xiong H, Gendelman HE. HIV-1-associated dementia: a metabolic encephalopathy perpetrated by virus-infected and immune-competent mononuclear phagocytes. *J Acquir Immune Defic Syndr*. 2002; 31 (Suppl 2):S43–54. [PubMed: 12394782]
4. Strizki JM, Albright AV, Sheng H, et al. Infection of primary human microglia and monocyte-derived macrophages with human immunodeficiency virus type 1 isolates: evidence of differential tropism. *J Virol*. 1996; 70:7654–62. [PubMed: 8892885]
5. Rubbert A, Combadiere C, Ostrowski M, et al. Dendritic cells express multiple chemokine receptors used as coreceptors for HIV entry. *J Immunol*. 1998; 160:3933–41. [PubMed: 9558100]
6. Clapham PR, McKnight A. HIV-1 receptors and cell tropism. *Br Med Bull*. 2001; 58:43–59. [PubMed: 11714623]
7. Ji H, Bracken C, Lu M. Buried polar interactions and conformational stability in the simian immunodeficiency virus (SIV) gp41 core. *Biochemistry*. 2000; 39:676–85. [PubMed: 10651632]
8. Liszewski MK, Yu JJ, O'Doherty U. Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR. *Methods*. 2009; 47:254–60. [PubMed: 19195495]
9. Simon-Loriere E, Rossolillo P, Negroni M. RNA structures, genomic organization and selection of recombinant HIV. *RNA Biol*. 2011; 8:280–6. [PubMed: 21422815]
10. Shah S, Nonnemacher MR, Pirrone V, Wigdahl B. Innate and adaptive factors regulating human immunodeficiency virus type 1 genomic activation. *J Neuroimmune Pharmacol*. 2010; 5:278–93. [PubMed: 20387125]
11. Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol*. 1997; 71:5382–90. [PubMed: 9188609]
12. Hansen MS, Bushman FD. Human immunodeficiency virus type 2 preintegration complexes: activities *in vitro* and response to inhibitors. *J Virol*. 1997; 71:3351–6. [PubMed: 9060709]
13. Bushman FD, Miller MD. Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J Virol*. 1997; 71:458–64. [PubMed: 8985371]
14. Cereseto A, Manganaro L, Gutierrez MI, et al. Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J*. 2005; 24:3070–81. [PubMed: 16096645]
15. Li L, Dahiya S, Kortagere S, et al. Impact of Tat Genetic Variation on HIV-1 Disease. *Advances in virology*. 2012; 2012:123605. [PubMed: 22899925]
16. Li L, Aiamkitsumrit B, Pirrone V, et al. Development of co-selected single nucleotide polymorphisms in the viral promoter precedes the onset of human immunodeficiency virus type 1-associated neurocognitive impairment. *J Neurovirol*. 2011; 17:92–109. [PubMed: 21225391]
17. Deng L, Wang D, de la Fuente C, et al. Enhancement of the p300 HAT activity by HIV-1 Tat on chromatin DNA. *Virology*. 2001; 289:312–26. [PubMed: 11689053]
18. McClure MO, Marsh M, Weiss RA. Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *EMBO J*. 1988; 7:513–8. [PubMed: 3259178]
19. Stein BS, Gowda SD, Lifson JD, et al. pH-independent HIV entry into CD4-positive T cells *via* virus envelope fusion to the plasma membrane. *Cell*. 1987; 49:659–68. [PubMed: 3107838]
20. Moore JP. Coreceptors: implications for HIV pathogenesis and therapy. *Science*. 1997; 276:51–2. [PubMed: 9122710]
21. Maddon PJ, Dalgleish AG, McDougal JS, et al. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*. 1986; 47:333–48. [PubMed: 3094962]
22. Oyaizu N, Chirmule N, Pahwa S. Role of CD4 molecule in the induction of interleukin 2 and interleukin 2 receptor in class II major histocompatibility complex-restricted antigen-specific T helper clones. T cell receptor/CD3 complex transmits CD4-dependent and CD4-independent signals. *J Clin Invest*. 1992; 89:1807–16. [PubMed: 1534818]
23. Zhang S, Zhang H, Zhao J. The role of CD4 T cell help for CD8 CTL activation. *Biochem Biophys Res Commun*. 2009; 384:405–8. [PubMed: 19410556]
24. Sattentau QJ, Weiss RA. The CD4 antigen: physiological ligand and HIV receptor. *Cell*. 1988; 52:631–3. [PubMed: 2830988]

25. Dalgleish AG, Beverley PC, Clapham PR, et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*. 1984; 312:763–7. [PubMed: 6096719]
26. Klatzmann D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*. 1984; 312:767–8. [PubMed: 6083454]
27. Klatzmann D, Barre-Sinoussi F, Nugeyre MT, et al. Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science*. 1984; 225:59–63. [PubMed: 6328660]
28. Guttman M, Kahn M, Garcia NK, Hu SL, Lee KK. Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. *J Virol*. 2012; 86:8750–64. [PubMed: 22674993]
29. Sattentau QJ, Moore JP. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J Exp Med*. 1991; 174:407–15. [PubMed: 1713252]
30. Lapham CK, Ouyang J, Chandrasekhar B, et al. Evidence for cell-surface association between fusin and the CD4-gp120 complex in human cell lines. *Science*. 1996; 274:602–5. [PubMed: 8849450]
31. Bleul CC, Farzan M, Choe H, et al. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature*. 1996; 382:829–33. [PubMed: 8752280]
32. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*. 1996; 272:872–7. [PubMed: 8629022]
33. Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1 α , MIP-1 β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*. 1996; 272:1955–8. [PubMed: 8658171]
34. Choe H, Farzan M, Sun Y, et al. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*. 1996; 85:1135–48. [PubMed: 8674119]
35. Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. 1996; 381:661–6. [PubMed: 8649511]
36. Dragic T, Litwin V, Allaway GP, et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*. 1996; 381:667–73. [PubMed: 8649512]
37. Doranz BJ, Rucker J, Yi Y, et al. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell*. 1996; 85:1149–58. [PubMed: 8674120]
38. Liao F, Alkhatib G, Peden KW, et al. STRL33, A novel chemokine receptor-like protein, functions as a fusion cofactor for both macrophage-tropic and T cell line-tropic HIV-1. *J Exp Med*. 1997; 185:2015–23. [PubMed: 9166430]
39. Choe H, Farzan M, Konkel M, et al. The orphan seven-transmembrane receptor apj supports the entry of primary T-cell-line-tropic and dualtropic human immunodeficiency virus type 1. *J Virol*. 1998; 72:6113–8. [PubMed: 9621075]
40. Allan JS, Coligan JE, Barin F, et al. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science*. 1985; 228:1091–4. [PubMed: 2986290]
41. Veronese FD, DeVico AL, Copeland TD, et al. Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science*. 1985; 229:1402–5. [PubMed: 2994223]
42. Leonard CK, Spellman MW, Riddle L, et al. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem*. 1990; 265:10373–82. [PubMed: 2355006]
43. Earl PL, Doms RW, Moss B. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc Natl Acad Sci U S A*. 1990; 87:648–52. [PubMed: 2300552]
44. Willey RL, Bonifacino JS, Potts BJ, Martin MA, Klausner RD. Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proc Natl Acad Sci U S A*. 1988; 85:9580–4. [PubMed: 2849111]

45. Decroly E, Wouters S, Di Bello C, et al. Identification of the paired basic convertases implicated in HIV gp160 processing based on *in vitro* assays and expression in CD4(+) cell lines. *J Biol Chem*. 1996; 271:30442–50. [PubMed: 8940009]
46. Helseth E, Olshevsky U, Furman C, Sodroski J. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J Virol*. 1991; 65:2119–23. [PubMed: 2002555]
47. Lasky LA, Groopman JE, Fennie CW, et al. Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. *Science*. 1986; 233:209–12. [PubMed: 3014647]
48. Collins-Fairclough AM, Charurat M, Nadai Y, et al. Significantly longer envelope V2 loops are characteristic of heterosexually transmitted subtype B HIV-1 in Trinidad. *PLoS One*. 2011; 6:e19995. [PubMed: 21698149]
49. Da LT, Quan JM, Wu YD. Understanding of the bridging sheet formation of HIV-1 glycoprotein gp120. *The journal of physical chemistry B*. 2009; 113:14536–43. [PubMed: 19813706]
50. McLellan JS, Pancera M, Carrico C, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*. 2011; 480:336–43. [PubMed: 22113616]
51. Saunders CJ, McCaffrey RA, Zharkikh I, et al. The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *J Virol*. 2005; 79:9069–80. [PubMed: 15994801]
52. Wan Y, Liu L, Wu L, et al. Deglycosylation or partial removal of HIV-1 CN54 gp140 V1/V2 domain enhances env-specific T cells. *AIDS Res Hum Retroviruses*. 2009; 25:607–17. [PubMed: 19500018]
53. Doria-Rose NA, Georgiev I, O'Dell S, et al. A short segment of the HIV-1 gp120 V1/V2 region is a major determinant of resistance to V1/V2 neutralizing antibodies. *J Virol*. 2012; 86:8319–23. [PubMed: 22623764]
54. Myers, G.; Korber, B.; Wain-Hobson, S.; Smith, RF.; Pavlakis, GN. *Theoretical Biology and Biophysics*. Los Alamos National Laboratory; Los Alamos, NM: 1993. Human retroviruses and AIDS 1993: a compilation and analysis of nucleic acid and amino acid sequences.
55. Wang WK, Essex M, Lee TH. The highly conserved aspartic acid residue between hypervariable regions 1 and 2 of human immunodeficiency virus type 1 gp120 is important for early stages of virus replication. *J Virol*. 1995; 69:538–42. [PubMed: 7983752]
56. Stamatos L, Wiskerchen M, Cheng-Mayer C. Effect of major deletions in the V1 and V2 loops of a macrophage-tropic HIV type 1 isolate on viral envelope structure, cell entry, and replication. *AIDS Res Hum Retroviruses*. 1998; 14:1129–39. [PubMed: 9737584]
57. Wyatt R, Sullivan N, Thali M, et al. Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J Virol*. 1993; 67:4557–65. [PubMed: 8331723]
58. Wyatt R, Moore J, Accola M, et al. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J Virol*. 1995; 69:5723–33. [PubMed: 7543586]
59. Bontjer I, Land A, Eggink D, et al. Optimization of human immunodeficiency virus type 1 envelope glycoproteins with V1/V2 deleted, using virus evolution. *J Virol*. 2009; 83:368–83. [PubMed: 18922866]
60. Cao J, Sullivan N, Desjardin E, et al. Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol*. 1997; 71:9808–12. [PubMed: 9371651]
61. Nabatov AA, Pollakis G, Linnemann T, et al. Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies. *J Virol*. 2004; 78:524–30. [PubMed: 14671134]
62. Rolland M, Edlefsen PT, Larsen BB, et al. Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature*. 2012; 490:417–20. [PubMed: 22960785]
63. Liu L, Cimbri R, Lusso P, Berger EA. Intraprotomer masking of third variable loop (V3) epitopes by the first and second variable loops (V1V2) within the native HIV-1 envelope glycoprotein trimer. *Proc Natl Acad Sci U S A*. 2011; 108:20148–53. [PubMed: 22128330]

64. Stamatatos L, Cheng-Mayer C. An envelope modification that renders a primary, neutralization-resistant clade B human immunodeficiency virus type 1 isolate highly susceptible to neutralization by sera from other clades. *J Virol.* 1998; 72:7840–5. [PubMed: 9733820]
65. Javaherian K, Langlois AJ, McDanal C, et al. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc Natl Acad Sci U S A.* 1989; 86:6768–72. [PubMed: 2771954]
66. Goudsmit J, Debouck C, Melen RH, et al. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. *Proc Natl Acad Sci U S A.* 1988; 85:4478–82. [PubMed: 2454471]
67. LaRosa GJ, Davide JP, Weinhold K, et al. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science.* 1990; 249:932–5. [PubMed: 2392685]
68. Sirois S, Sing T, Chou KC. HIV-1 gp120 V3 loop for structure-based drug design. *Curr Protein Pept Sci.* 2005; 6:413–22. [PubMed: 16248793]
69. Sharon M, Kessler N, Levy R, et al. Alternative conformations of HIV-1 V3 loops mimic beta hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure.* 2003; 11:225–36. [PubMed: 12575942]
70. Korber BT, Farber RM, Wolpert DH, Lapedes AS. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc Natl Acad Sci U S A.* 1993; 90:7176–80. [PubMed: 8346232]
71. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J. 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; 66:6777–80. [PubMed: 1404617]
72. Cormier EG, Dragic T. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. *J Virol.* 2002; 76:8953–7. [PubMed: 12163614]
73. 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; 66:3183–7. [PubMed: 1560543]
74. Briggs DR, Tuttle DL, Sleasman JW, Goodenow MM. Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). *AIDS.* 2000; 14:2937–9. [PubMed: 11153675]
75. Xiao L, Owen SM, Goldman I, et al. CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology.* 1998; 240:83–92. [PubMed: 9448692]
76. Huang CC, Tang M, Zhang MY, et al. Structure of a V3-containing HIV-1 gp120 core. *Science.* 2005; 310:1025–8. [PubMed: 16284180]
77. Dobrowsky TM, Zhou Y, Sun SX, Siliciano RF, Wirtz D. Monitoring early fusion dynamics of human immunodeficiency virus type 1 at single-molecule resolution. *J Virol.* 2008; 82:7022–33. [PubMed: 18480458]
78. Dimitrov DS, Xiao X, Chabot DJ, Broder CC. HIV coreceptors. *J Membr Biol.* 1998; 166:75–90. [PubMed: 9841733]
79. Douglas NW, Munro GH, Daniels RS. HIV/SIV glycoproteins: structure-function relationships. *J Mol Biol.* 1997; 273:122–49. [PubMed: 9367752]
80. Kinsey NE, Anderson MG, Unangst TJ, et al. Antigenic variation of SIV: mutations in V4 alter the neutralization profile. *Virology.* 1996; 221:14–21. [PubMed: 8661410]
81. Overbaugh J, Rudensey LM. Alterations in potential sites for glycosylation predominate during evolution of the simian immunodeficiency virus envelope gene in macaques. *J Virol.* 1992; 66:5937–48. [PubMed: 1527847]
82. Torres JV, Malley A, Banapour B, et al. An epitope on the surface envelope glycoprotein (gp130) of simian immunodeficiency virus (SIVmac) involved in viral neutralization and T cell activation. *AIDS Res Hum Retroviruses.* 1993; 9:423–30. [PubMed: 7686386]
83. Karray S, Zouali M. Identification of the B cell superantigen-binding site of HIV-1 gp120. *Proc Natl Acad Sci U S A.* 1997; 94:1356–60. [PubMed: 9037057]

84. Berberian L, Shukla J, Jefferis R, Braun J. Effects of HIV infection on VH3 (D12 idiotope) B cells *in vivo*. J Acquir Immune Defic Syndr. 1994; 7:641–6. [PubMed: 8207642]
85. Zouali M. B-cell superantigens: implications for selection of the human antibody repertoire. Immunol Today. 1995; 16:399–405. [PubMed: 7546197]
86. Kwong PD, Wyatt R, Robinson J, et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998; 393:648–59. [PubMed: 9641677]
87. Sanders RW, van Anken E, Nabatov AA, et al. The carbohydrate at asparagine 386 on HIV-1 gp120 is not essential for protein folding and function but is involved in immune evasion. Retrovirology. 2008; 5:10. [PubMed: 18237398]
88. Edwards TG, Hoffman TL, Baribaud F, et al. Relationships between CD4 independence, neutralization sensitivity, and exposure of a CD4-induced epitope in a human immunodeficiency virus type 1 envelope protein. J Virol. 2001; 75:5230–9. [PubMed: 11333905]
89. Chamberland A, Sylla M, Boulassel MR, et al. Effect of antiretroviral therapy on HIV-1 genetic evolution during acute infection. International journal of STD & AIDS. 2011; 22:146–50. [PubMed: 21464451]
90. Cheng-Mayer C, Brown A, Harouse J, Luciw PA, Mayer AJ. Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant *in vivo* by virtue of sequence changes in the extracellular envelope glycoprotein that modify N-linked glycosylation. J Virol. 1999; 73:5294–300. [PubMed: 10364275]
91. Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1. Nature. 2003; 422:307–12. [PubMed: 12646921]
92. Sagar M, Wu X, Lee S, Overbaugh J. Human immunodeficiency virus type 1 V1–V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. J Virol. 2006; 80:9586–98. [PubMed: 16973562]
93. Wang W, Nie J, Prochnow C, et al. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. Retrovirology. 2013; 10:14. [PubMed: 23384254]
94. Narayan KM, Agrawal N, Du SX, et al. Prime-boost immunization of rabbits with HIV-1 gp120 elicits potent neutralization activity against a primary viral isolate. PLoS One. 2013; 8:e52732. [PubMed: 23326351]
95. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. Nature. 1997; 387:426–30. [PubMed: 9163431]
96. Gallaher WR. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell. 1987; 50:327–8. [PubMed: 3496970]
97. Freed EO, Myers DJ, Risser R. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. Proc Natl Acad Sci U S A. 1990; 87:4650–4. [PubMed: 2191297]
98. Chambers P, Pringle CR, Easton AJ. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. J Gen Virol. 1990; 71 (Pt 12):3075–80. [PubMed: 2177097]
99. Delwart EL, Mosialos G, Gilmore T. Retroviral envelope glycoproteins contain a “leucine zipper”-like repeat. AIDS Res Hum Retroviruses. 1990; 6:703–6. [PubMed: 2364015]
100. O’Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science. 1991; 254:539–44. [PubMed: 1948029]
101. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science. 1988; 240:1759–64. [PubMed: 3289117]
102. Dubay JW, Roberts SJ, Brody B, Hunter E. Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. J Virol. 1992; 66:4748–56. [PubMed: 1629954]
103. Pombourios P, el Ahmar W, McPhee DA, Kemp BE. Determinants of human immunodeficiency virus type 1 envelope glycoprotein oligomeric structure. J Virol. 1995; 69:1209–18. [PubMed: 7815497]

104. Shang L, Yue L, Hunter E. Role of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein in cell-cell fusion and virus infection. *J Virol.* 2008; 82:5417–28. [PubMed: 18353944]
105. Welman M, Lemay G, Cohen EA. Role of envelope processing and gp41 membrane spanning domain in the formation of human immunodeficiency virus type 1 (HIV-1) fusion-competent envelope glycoprotein complex. *Virus Res.* 2007; 124:103–12. [PubMed: 17129629]
106. Kowalski M, Potz J, Basiripour L, et al. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science.* 1987; 237:1351–5. [PubMed: 3629244]
107. Yue L, Shang L, Hunter E. Truncation of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein defines elements required for fusion, incorporation, and infectivity. *J Virol.* 2009; 83:11588–98. [PubMed: 19726514]
108. Shang L, Hunter E. Residues in the membrane-spanning domain core modulate conformation and fusogenicity of the HIV-1 envelope glycoprotein. *Virology.* 2010; 404:158–67. [PubMed: 20605619]
109. Ratner L, Fisher A, Jagodzinski LL, et al. Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res Hum Retroviruses.* 1987; 3:57–69. [PubMed: 3040055]
110. Sonigo P, Alizon M, Staskus K, et al. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell.* 1985; 42:369–82. [PubMed: 2410140]
111. Owens RJ, Burke C, Rose JK. Mutations in the membrane-spanning domain of the human immunodeficiency virus envelope glycoprotein that affect fusion activity. *J Virol.* 1994; 68:570–4. [PubMed: 8254774]
112. Gabuzda DH, Lever A, Terwilliger E, Sodroski J. Effects of deletions in the cytoplasmic domain on biological functions of human immunodeficiency virus type 1 envelope glycoproteins. *J Virol.* 1992; 66:3306–15. [PubMed: 1583717]
113. Murakami T, Freed EO. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc Natl Acad Sci U S A.* 2000; 97:343–8. [PubMed: 10618420]
114. Yu X, Yuan X, McLane MF, Lee TH, Essex M. Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *J Virol.* 1993; 67:213–21. [PubMed: 8416370]
115. Lee SJ, Hu W, Fisher AG, et al. Role of the carboxy-terminal portion of the HIV-1 transmembrane protein in viral transmission and cytopathogenicity. *AIDS Res Hum Retroviruses.* 1989; 5:441–9. [PubMed: 2788444]
116. Dubay JW, Roberts SJ, Hahn BH, Hunter E. Truncation of the human immunodeficiency virus type 1 transmembrane glycoprotein cytoplasmic domain blocks virus infectivity. *J Virol.* 1992; 66:6616–25. [PubMed: 1357190]
117. Wilk T, Pfeiffer T, Bosch V. Retained *in vitro* infectivity and cytopathogenicity of HIV-1 despite truncation of the C-terminal tail of the env gene product. *Virology.* 1992; 189:167–77. [PubMed: 1604808]
118. Clapham PR, Blanc D, Weiss RA. Specific cell surface requirements for the infection of CD4-positive cells by human immunodeficiency virus types 1 and 2 and by Simian immunodeficiency virus. *Virology.* 1991; 181:703–15. [PubMed: 1673040]
119. Choi WT, An J. Biology and clinical relevance of chemokines and chemokine receptors CXCR4 and CCR5 in human diseases. *Exp Biol Med (Maywood).* 2011; 236:637–47. [PubMed: 21565895]
120. Fauci AS. Host factors and the pathogenesis of HIV-induced disease. *Nature.* 1996; 384:529–34. [PubMed: 8955267]
121. Jinno A, Shimizu N, Soda Y, et al. Identification of the chemokine receptor TER1/CCR8 expressed in brain-derived cells and T cells as a new coreceptor for HIV-1 infection. *Biochem Biophys Res Commun.* 1998; 243:497–502. [PubMed: 9480837]
122. Murphy PM. The molecular biology of leukocyte chemoattractant receptors. *Annu Rev Immunol.* 1994; 12:593–633. [PubMed: 8011292]
123. Baggiolini M, Dewald B, Moser B. Human chemokines: an update. *Annu Rev Immunol.* 1997; 15:675–705. [PubMed: 9143704]

124. Wu B, Chien EY, Mol CD, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*. 2010; 330:1066–71. [PubMed: 20929726]
125. Bleul CC, Wu L, Hoxie JA, Springer TA, Mackay CR. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci U S A*. 1997; 94:1925–30. [PubMed: 9050881]
126. Albright AV, Shieh JT, Itoh T, et al. Microglia express CCR5, CXCR4, and CCR3, but of these, CCR5 is the principal coreceptor for human immunodeficiency virus type 1 dementia isolates. *J Virol*. 1999; 73:205–13. [PubMed: 9847323]
127. Reckamp KL, Figlin RA, Burdick MD, et al. CXCR4 expression on circulating pan-cytokeratin positive cells is associated with survival in patients with advanced non-small cell lung cancer. *BMC Cancer*. 2009; 9:213. [PubMed: 19563666]
128. Alkhatib G. The biology of CCR5 and CXCR4. *Curr Opin HIV AIDS*. 2009; 4:96–103. [PubMed: 19339947]
129. Haribabu B, Richardson RM, Fisher I, et al. Regulation of human chemokine receptors CXCR4. Role of phosphorylation in desensitization and internalization. *J Biol Chem*. 1997; 272:28726–31. [PubMed: 9353342]
130. Ganju RK, Brubaker SA, Meyer J, et al. The alpha-chemokine, stromal cell-derived factor-1alpha, binds to the transmembrane G-protein-coupled CXCR-4 receptor and activates multiple signal transduction pathways. *J Biol Chem*. 1998; 273:23169–75. [PubMed: 9722546]
131. Kryczek I, Wei S, Keller E, Liu R, Zou W. Stroma-derived factor (SDF-1/CXCL12) and human tumor pathogenesis. *Am J Physiol Cell Physiol*. 2007; 292:C987–95. [PubMed: 16943240]
132. Zhang XF, Wang JF, Matczak E, Proper JA, Groopman JE. Janus kinase 2 is involved in stromal cell-derived factor-1alpha-induced tyrosine phosphorylation of focal adhesion proteins and migration of hematopoietic progenitor cells. *Blood*. 2001; 97:3342–8. [PubMed: 11369622]
133. Zou YR, Kottmann AH, Kuroda M, Taniuchi I, Littman DR. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature*. 1998; 393:595–9. [PubMed: 9634238]
134. Ma Q, Jones D, Borghesani PR, et al. Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice. *Proc Natl Acad Sci U S A*. 1998; 95:9448–53. [PubMed: 9689100]
135. Picard L, Wilkinson DA, McKnight A, et al. Role of the aminoterminal extracellular domain of CXCR-4 in human immunodeficiency virus type 1 entry. *Virology*. 1997; 231:105–11. [PubMed: 9143308]
136. Brelot A, Heveker N, Pleskoff O, Sol N, Alizon M. Role of the first and third extracellular domains of CXCR-4 in human immunodeficiency virus coreceptor activity. *J Virol*. 1997; 71:4744–51. [PubMed: 9151868]
137. Brelot A, Heveker N, Montes M, Alizon M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem*. 2000; 275:23736–44. [PubMed: 10825158]
138. Oberlin E, Amara A, Bachelier F, et al. The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature*. 1996; 382:833–5. [PubMed: 8752281]
139. Winkler C, Modi W, Smith MW, et al. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science*. 1998; 279:389–93. [PubMed: 9430590]
140. Doranz BJ, Orsini MJ, Turner JD, et al. Identification of CXCR4 domains that support coreceptor and chemokine receptor functions. *J Virol*. 1999; 73:2752–61. [PubMed: 10074122]
141. Zhou N, Luo Z, Luo J, et al. Structural and functional characterization of human CXCR4 as a chemokine receptor and HIV-1 co-receptor by mutagenesis and molecular modeling studies. *J Biol Chem*. 2001; 276:42826–33. [PubMed: 11551942]
142. Tian S, Choi WT, Liu D, et al. Distinct functional sites for human immunodeficiency virus type 1 and stromal cell-derived factor 1alpha on CXCR4 transmembrane helical domains. *J Virol*. 2005; 79:12667–73. [PubMed: 16188969]

143. Wang ZX, Berson JF, Zhang TY, et al. CXCR4 sequences involved in coreceptor determination of human immunodeficiency virus type-1 tropism. Unmasking of activity with M-tropic Env glycoproteins. *J Biol Chem.* 1998; 273:15007–15. [PubMed: 9614108]
144. Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry.* 1996; 35:3362–7. [PubMed: 8639485]
145. Blanpain C, Migeotte I, Lee B, et al. CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. *Blood.* 1999; 94:1899–905. [PubMed: 10477718]
146. Wells TN, Power CA, Lusti-Narasimhan M, et al. Selectivity and antagonism of chemokine receptors. *J Leukoc Biol.* 1996; 59:53–60. [PubMed: 8558067]
147. Levy JP. Questions about CD8+ anti-HIV lymphocytes in the control of HIV infection. *Antibiot Chemother.* 1996; 48:13–20. [PubMed: 8726501]
148. Cocchi F, DeVico AL, Garzino-Demo A, et al. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science.* 1995; 270:1811–5. [PubMed: 8525373]
149. Rucker J, Samson M, Doranz BJ, et al. Regions in beta-chemokine receptors CCR5 and CCR2b that determine HIV-1 cofactor specificity. *Cell.* 1996; 87:437–46. [PubMed: 8898197]
150. Balistreri CR, Caruso C, Grimaldi MP, et al. CCR5 receptor: biologic and genetic implications in age-related diseases. *Ann N Y Acad Sci.* 2007; 1100:162–72. [PubMed: 17460174]
151. Sorce S, Myburgh R, Krause KH. The chemokine receptor CCR5 in the central nervous system. *Prog Neurobiol.* 2011; 93:297–311. [PubMed: 21163326]
152. Shieh JT, Albright AV, Sharron M, et al. Chemokine receptor utilization by human immunodeficiency virus type 1 isolates that replicate in microglia. *J Virol.* 1998; 72:4243–9. [PubMed: 9557714]
153. Dorf ME, Berman MA, Tanabe S, Heesen M, Luo Y. Astrocytes express functional chemokine receptors. *J Neuroimmunol.* 2000; 111:109–21. [PubMed: 11063828]
154. Martin-Garcia J, Kolson DL, Gonzalez-Scarano F. Chemokine receptors in the brain: their role in HIV infection and pathogenesis. *AIDS.* 2002; 16:1709–30. [PubMed: 12218382]
155. O'Brien TR, Winkler C, Dean M, et al. HIV-1 infection in a man homozygous for CCR5 delta 32. *Lancet.* 1997; 349:1219. [PubMed: 9130945]
156. Libert F, Cochaux P, Beckman G, et al. The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet.* 1998; 7:399–406. [PubMed: 9466996]
157. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet.* 1997; 16:100–3. [PubMed: 9140404]
158. Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell.* 1996; 86:367–77. [PubMed: 8756719]
159. Michael NL, Nelson JA, KewalRamani VN, et al. Exclusive and persistent use of the entry coreceptor CXCR4 by human immunodeficiency virus type 1 from a subject homozygous for CCR5 delta32. *J Virol.* 1998; 72:6040–7. [PubMed: 9621067]
160. Agrawal L, Lu X, Qingwen J, et al. Role for CCR5Delta32 protein in resistance to R5, R5X4, and X4 human immunodeficiency virus type 1 in primary CD4+ cells. *J Virol.* 2004; 78:2277–87. [PubMed: 14963124]
161. Jin Q, Marsh J, Cornetta K, Alkhatib G. Resistance to human immunodeficiency virus type 1 (HIV-1) generated by lentivirus vector-mediated delivery of the CCR5{Delta}32 gene despite detectable expression of the HIV-1 co-receptors. *J Gen Virol.* 2008; 89:2611–21. [PubMed: 18796731]
162. Bajetto A, Bonavia R, Barbero S, Florio T, Schettini G. Chemokines and their receptors in the central nervous system. *Front Neuroendocrinol.* 2001; 22:147–84. [PubMed: 11456467]
163. Horuk R, Martin AW, Wang Z, et al. Expression of chemokine receptors by subsets of neurons in the central nervous system. *J Immunol.* 1997; 158:2882–90. [PubMed: 9058825]
164. Xia MQ, Bacskai BJ, Knowles RB, Qin SX, Hyman BT. Expression of the chemokine receptor CXCR3 on neurons and the elevated expression of its ligand IP-10 in reactive astrocytes: *in vitro*

- ERK1/2 activation and role in Alzheimer's disease. *J Neuroimmunol.* 2000; 108:227–35. [PubMed: 10900358]
165. He J, Chen Y, Farzan M, et al. CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature.* 1997; 385:645–9. [PubMed: 9024664]
 166. Gupta SK, Lysko PG, Pillarisetti K, Ohlstein E, Stadel JM. Chemokine receptors in human endothelial cells. Functional expression of CXCR4 and its transcriptional regulation by inflammatory cytokines. *J Biol Chem.* 1998; 273:4282–7. [PubMed: 9461627]
 167. Sinclair E, Gray F, Ciardi A, Scaravilli F. Immunohistochemical changes and PCR detection of HIV provirus DNA in brains of asymptomatic HIV-positive patients. *J Neuropathol Exp Neurol.* 1994; 53:43–50. [PubMed: 8301319]
 168. Koenig S, Gendelman HE, Orenstein JM, et al. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science.* 1986; 233:1089–93. [PubMed: 3016903]
 169. Gray L, Roche M, Churchill MJ, et al. Tissue-specific sequence alterations in the human immunodeficiency virus type 1 envelope favoring CCR5 usage contribute to persistence of dual-tropic virus in the brain. *J Virol.* 2009; 83:5430–41. [PubMed: 19321618]
 170. Ricardo-Dukelow M, Kadiu I, Rozek W, et al. HIV-1 infected monocyte-derived macrophages affect the human brain microvascular endothelial cell proteome: new insights into blood-brain barrier dysfunction for HIV-1-associated dementia. *J Neuroimmunol.* 2007; 185:37–46. [PubMed: 17321604]
 171. Martin-Garcia J, Cao W, Varela-Rohena A, Plassmeyer ML, Gonzalez-Scarano F. HIV-1 tropism for the central nervous system: Brain-derived envelope glycoproteins with lower CD4 dependence and reduced sensitivity to a fusion inhibitor. *Virology.* 2006; 346:169–79. [PubMed: 16309726]
 172. Price RW, Brew B, Sidtis J, et al. The brain in AIDS: central nervous system HIV-1 infection and AIDS dementia complex. *Science.* 1988; 239:586–92. [PubMed: 3277272]
 173. Vallat AV, De Girolami U, He J, et al. Localization of HIV-1 co-receptors CCR5 and CXCR4 in the brain of children with AIDS. *Am J Pathol.* 1998; 152:167–78. [PubMed: 9422534]
 174. Sanders VJ, Pittman CA, White MG, et al. Chemokines and receptors in HIV encephalitis. *AIDS.* 1998; 12:1021–6. [PubMed: 9662198]
 175. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol.* 1999; 17:657–700. [PubMed: 10358771]
 176. Rucker J, Edinger AL, Sharron M, et al. Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses. *J Virol.* 1997; 71:8999–9007. [PubMed: 9371556]
 177. Bjorndal A, Deng H, Jansson M, et al. Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol.* 1997; 71:7478–87. [PubMed: 9311827]
 178. Zhang L, He T, Huang Y, et al. Chemokine coreceptor usage by diverse primary isolates of human immunodeficiency virus type 1. *J Virol.* 1998; 72:9307–12. [PubMed: 9765480]
 179. Thielen A, Sichtig N, Kaiser R, et al. Improved prediction of HIV-1 coreceptor usage with sequence information from the second hypervariable loop of gp120. *J Infect Dis.* 2010; 202:1435–43. [PubMed: 20874088]
 180. Moyle GJ, Wildfire A, Mandalia S, et al. Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection. *J Infect Dis.* 2005; 191:866–72. [PubMed: 15717260]
 181. Wang T, Xu Y, Zhu H, et al. Successful isolation of infectious and high titer human monocyte-derived HIV-1 from two subjects with discontinued therapy. *PLoS One.* 2013; 8:e65071. [PubMed: 23741458]
 182. Fischer-Smith T, Croul S, Sverstiuk AE, et al. CNS invasion by CD14+/CD16+ peripheral blood-derived monocytes in HIV dementia: perivascular accumulation and reservoir of HIV infection. *J Neurovirol.* 2001; 7:528–41. [PubMed: 11704885]
 183. Gendelman HE, Orenstein JM, Martin MA, et al. Efficient isolation and propagation of human immunodeficiency virus on recombinant colony-stimulating factor 1-treated monocytes. *J Exp Med.* 1988; 167:1428–41. [PubMed: 3258626]

184. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol.* 2005; 5:953–64. [PubMed: 16322748]
185. Swingle S, Mann AM, Zhou J, Swingle C, Stevenson M. Apoptotic killing of HIV-1-infected macrophages is subverted by the viral envelope glycoprotein. *PLoS Pathog.* 2007; 3:1281–90. [PubMed: 17907802]
186. Benaroch P, Billard E, Gaudin R, Schindler M, Jouve M. HIV-1 assembly in macrophages. *Retrovirology.* 2010; 7:29. [PubMed: 20374631]
187. Koppensteiner H, Banning C, Schneider C, Hohenberg H, Schindler M. Macrophage internal HIV-1 is protected from neutralizing antibodies. *J Virol.* 2012; 86:2826–36. [PubMed: 22205742]
188. Gonzalez-Scarano F, Martin-Garcia J. The neuropathogenesis of AIDS. *Nat Rev Immunol.* 2005; 5:69–81. [PubMed: 15630430]
189. Joseph SB, Arrildt KT, Swanstrom AE, et al. Quantification of Entry Phenotypes of Macrophage-Tropic HIV-1 across a Wide Range of CD4 Densities. *J Virol.* 2014; 88:1858–69. [PubMed: 24307580]
190. Davenport MP, Zaunders JJ, Hazenberg MD, Schuitemaker H, van Rij RP. Cell turnover and cell tropism in HIV-1 infection. *Trends in microbiology.* 2002; 10:275–8. [PubMed: 12088663]
191. Wightman F, Solomon A, Khoury G, et al. Both CD31(+) and CD31(–) naive CD4(+) T cells are persistent HIV type 1-infected reservoirs in individuals receiving antiretroviral therapy. *J Infect Dis.* 2010; 202:1738–48. [PubMed: 20979453]
192. Stevenson M. HIV-1 pathogenesis. *Nat Med.* 2003; 9:853–60. [PubMed: 12835705]
193. Pan X, Baldauf HM, Keppler OT, Fackler OT. Restrictions to HIV-1 replication in resting CD4+ T lymphocytes. *Cell research.* 2013; 23:876–85. [PubMed: 23732522]
194. Cameron PU, Saleh S, Sallmann G, et al. Establishment of HIV-1 latency in resting CD4+ T cells depends on chemokine-induced changes in the actin cytoskeleton. *Proc Natl Acad Sci U S A.* 2010; 107:16934–9. [PubMed: 20837531]
195. Chege D, Sheth PM, Kain T, et al. Sigmoid Th17 populations, the HIV latent reservoir, and microbial translocation in men on long-term antiretroviral therapy. *AIDS.* 2011; 25:741–9. [PubMed: 21378536]
196. Eisele E, Siliciano RF. Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity.* 2012; 37:377–88. [PubMed: 22999944]
197. Gougeon ML, Lecoer H, Dulioust A, et al. Programmed cell death in peripheral lymphocytes from HIV-infected persons: increased susceptibility to apoptosis of CD4 and CD8 T cells correlates with lymphocyte activation and with disease progression. *J Immunol.* 1996; 156:3509–20. [PubMed: 8617980]
198. Grivel JC, Malkevitch N, Margolis L. Human immunodeficiency virus type 1 induces apoptosis in CD4(+) but not in CD8(+) T cells in ex vivo-infected human lymphoid tissue. *J Virol.* 2000; 74:8077–84. [PubMed: 10933717]
199. Doitsh G, Galloway NL, Geng X, et al. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature.* 2014; 505:509–14. [PubMed: 24356306]
200. Blaak H, van't Wout AB, Brouwer M, et al. *In vivo* HIV-1 infection of CD45RA(+)CD4(+) T cells is established primarily by syncytium-inducing variants and correlates with the rate of CD4(+) T cell decline. *Proc Natl Acad Sci U S A.* 2000; 97:1269–74. [PubMed: 10655520]
201. Chomont N, El-Far M, Ancuta P, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine.* 2009; 15:893–900.
202. Ochsenbauer C, Edmonds TG, Ding H, et al. Generation of transmitted/founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages. *J Virol.* 2012; 86:2715–28. [PubMed: 22190722]
203. Ping LH, Joseph SB, Anderson JA, et al. Comparison of viral Env proteins from acute and chronic infections with subtype C human immunodeficiency virus type 1 identifies differences in glycosylation and CCR5 utilization and suggests a new strategy for immunogen design. *J Virol.* 2013; 87:7218–33. [PubMed: 23616655]
204. Parrish NF, Gao F, Li H, et al. Phenotypic properties of transmitted founder HIV-1. *Proc Natl Acad Sci U S A.* 2013; 110:6626–33. [PubMed: 23542380]

205. Dejudicq N, Simmons G, Clapham PR. Expanded tropism of primary human immunodeficiency virus type 1 R5 strains to CD4(+) T-cell lines determined by the capacity to exploit low concentrations of CCR5. *J Virol.* 1999; 73:7842–7. [PubMed: 10438877]
206. Peters PJ, Sullivan WM, Duenas-Decamp MJ, et al. Non-macrophage-tropic human immunodeficiency virus type 1 R5 envelopes predominate in blood, lymph nodes, and semen: implications for transmission and pathogenesis. *J Virol.* 2006; 80:6324–32. [PubMed: 16775320]
207. Peters PJ, Bhattacharya J, Hibbitts S, et al. Biological analysis of human immunodeficiency virus type 1 R5 envelopes amplified from brain and lymph node tissues of AIDS patients with neuropathology reveals two distinct tropism phenotypes and identifies envelopes in the brain that confer an enhanced tropism and fusogenicity for macrophages. *J Virol.* 2004; 78:6915–26. [PubMed: 15194768]
208. Gartner S, Markovits P, Markovitz DM, et al. The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science.* 1986; 233:215–9. [PubMed: 3014648]
209. Peters PJ, Duenas-Decamp MJ, Sullivan WM, Clapham PR. Variation of macrophage tropism among HIV-1 R5 envelopes in brain and other tissues. *J Neuroimmune Pharmacol.* 2007; 2:32–41. [PubMed: 18040824]
210. Gorry PR, Bristol G, Zack JA, et al. Macrophage tropism of human immunodeficiency virus type 1 isolates from brain and lymphoid tissues predicts neurotropism independent of coreceptor specificity. *J Virol.* 2001; 75:10073–89. [PubMed: 11581376]
211. Gonzalez-Perez MP, O'Connell O, Lin R, et al. Independent evolution of macrophage-tropism and increased charge between HIV-1 R5 envelopes present in brain and immune tissue. *Retrovirology.* 2012; 9:20. [PubMed: 22420378]
212. Holman AG, Mefford ME, O'Connor N, Gabuzda D. HIVBrainSeqDB: a database of annotated HIV envelope sequences from brain and other anatomical sites. *AIDS research and therapy.* 2010; 7:43. [PubMed: 21156070]
213. Arrildt KT, Joseph SB, Swanson R. The HIV-1 env protein: a coat of many colors. *Current HIV/AIDS reports.* 2012; 9:52–63. [PubMed: 22237899]
214. Zamarchi R, Allavena P, Borsetti A, et al. Expression and functional activity of CXCR-4 and CCR-5 chemokine receptors in human thymocytes. *Clin Exp Immunol.* 2002; 127:321–30. [PubMed: 11876757]
215. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol.* 1999; 73:10489–502. [PubMed: 10559367]
216. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell lifespan, and viral generation time. *Science.* 1996; 271:1582–6. [PubMed: 8599114]
217. Le Douce V, Herbein G, Rohr O, Schwartz C. Molecular mechanisms of HIV-1 persistence in the monocyte-macrophage lineage. *Retrovirology.* 2010; 7:32. [PubMed: 20380694]
218. Lambotte O, Taoufik Y, de Goer MG, et al. Detection of infectious HIV in circulating monocytes from patients on prolonged highly active antiretroviral therapy. *Journal of acquired immune deficiency syndromes.* 2000; 23:114–9. [PubMed: 10737425]
219. Zhu T. HIV-1 genotypes in peripheral blood monocytes. *Journal of leukocyte biology.* 2000; 68:338–44. [PubMed: 10985249]
220. Zhu T, Muthui D, Holte S, et al. Evidence for human immunodeficiency virus type 1 replication *in vivo* in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *Journal of virology.* 2002; 76:707–16. [PubMed: 11752161]
221. Ellery PJ, Tippet E, Chiu YL, et al. The CD16+ monocyte subset is more permissive to infection and preferentially harbors HIV-1 *in vivo*. *J Immunol.* 2007; 178:6581–9. [PubMed: 17475889]
222. Koppensteiner H, Brack-Werner R, Schindler M. Macrophages and their relevance in Human Immunodeficiency Virus Type I infection. *Retrovirology.* 2012; 9:82. [PubMed: 23035819]
223. Strazza M, Pirrone V, Wigdahl B, Nonnemacher MR. Breaking down the barrier: the effects of HIV-1 on the blood-brain barrier. *Brain Res.* 2011; 1399:96–115. [PubMed: 21641584]

224. Crowe S, Zhu T, Muller WA. The contribution of monocyte infection and trafficking to viral persistence, and maintenance of the viral reservoir in HIV infection. *Journal of leukocyte biology*. 2003; 74:635–41. [PubMed: 12960232]
225. Verhofstede C, Brudney D, Reynaerts J, et al. Concordance between HIV-1 genotypic coreceptor tropism predictions based on plasma RNA and proviral DNA. *HIV Med*. 2011; 12:544–52. [PubMed: 21518222]
226. Coakley E, Petropoulos CJ, Whitcomb JM. Assessing chemokine co-receptor usage in HIV. *Curr Opin Infect Dis*. 2005; 18:9–15. [PubMed: 15647694]
227. Skrabal K, Low AJ, Dong W, et al. Determining human immunodeficiency virus coreceptor use in a clinical setting: degree of correlation between two phenotypic assays and a bioinformatic model. *J Clin Microbiol*. 2007; 45:279–84. [PubMed: 17122004]
228. Damer M, Kaiser R, Klein R, et al. Genotypic tropism testing by massively parallel sequencing: qualitative and quantitative analysis. *BMC Med Inform Decis Mak*. 2011; 11:30. [PubMed: 21569501]
229. Sierra S, Kaiser R, Thielen A, Lengauer T. Genotypic coreceptor analysis. *Eur J Med Res*. 2007; 12:453–62. [PubMed: 17933727]
230. McGovern RA, Thielen A, Mo T, et al. Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS*. 2010; 24:2517–25. [PubMed: 20736814]
231. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol*. 2007; 25:1407–10. [PubMed: 18066037]
232. Sing T, Low AJ, Beerenwinkel N, et al. Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther*. 2007; 12:1097–106. [PubMed: 18018768]
233. Low AJ, Swenson LC, Harrigan PR. HIV coreceptor phenotyping in the clinical setting. *AIDS Rev*. 2008; 10:143–51. [PubMed: 18820716]
234. Low AJ, Dong W, Chan D, et al. Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. *Aids*. 2007; 21:F17–24. [PubMed: 17721088]
235. Vandekerckhove LP, Wensing AM, Kaiser R, et al. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis*. 2011; 11:394–407. [PubMed: 21429803]
236. Leitner T, Halapi E, Scarlatti G, et al. Analysis of heterogeneous viral populations by direct DNA sequencing. *Biotechniques*. 1993; 15:120–7. [PubMed: 8363827]
237. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437:376–80. [PubMed: 16056220]
238. Swenson LC, Mo T, Dong WW, et al. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J Infect Dis*. 2011; 203:237–45. [PubMed: 21288824]
239. Suphaphiphat P, Essex M, Lee TH. Mutations in the V3 stem versus the V3 crown and C4 region have different effects on the binding and fusion steps of human immunodeficiency virus type 1 gp120 interaction with the CCR5 coreceptor. *Virology*. 2007; 360:182–90. [PubMed: 17101166]
240. Huang W, Toma J, Fransen S, et al. Coreceptor tropism can be influenced by amino acid substitutions in the gp41 transmembrane subunit of human immunodeficiency virus type 1 envelope protein. *J Virol*. 2008; 82:5584–93. [PubMed: 18353956]
241. Carrillo A, Ratner L. Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells. *J Virol*. 1996; 70:1310–6. [PubMed: 8551601]
242. Labrosse B, Treboute C, Brelot A, Alizon M. Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor. *J Virol*. 2001; 75:5457–64. [PubMed: 11356952]
243. Milich L, Margolin B, Swanstrom R. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. *J Virol*. 1993; 67:5623–34. [PubMed: 8350415]
244. Liu S, Fan S, Sun Z. Structural and functional characterization of the human CCR5 receptor in complex with HIV gp120 envelope glycoprotein and CD4 receptor by molecular modeling studies. *J Mol Model*. 2003; 9:329–36. [PubMed: 14517611]

245. Moore JP, Stevenson M. New targets for inhibitors of HIV-1 replication. *Nat Rev Mol Cell Biol.* 2000; 1:40–9. [PubMed: 11413488]
246. Poveda E, Alcamí J, Paredes R, et al. Genotypic determination of HIV tropism - clinical and methodological recommendations to guide the therapeutic use of CCR5 antagonists. *AIDS Rev.* 2010; 12:135–48. [PubMed: 20842202]
247. Raymond S, Delobel P, Mavigner M, et al. Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *AIDS.* 2008; 22:F11–6. [PubMed: 18753930]
248. Pillai S, Good B, Richman D, Corbeil J. A new perspective on V3 phenotype prediction. *AIDS Res Hum Retroviruses.* 2003; 19:145–9. [PubMed: 12643277]
249. Rizzuto CD, Wyatt R, Hernandez-Ramos N, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science.* 1998; 280:1949–53. [PubMed: 9632396]
250. Jensen, M. PSSM 2007–2009. Available from: <http://indra.mullins.microbiol.washington.edu/webpssm/>
251. Jensen MA, Li FS, van 't Wout AB, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol.* 2003; 77:13376–88. [PubMed: 14645592]
252. Resch W, Hoffman N, Swanstrom R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology.* 2001; 288:51–62. [PubMed: 11543657]
253. Soulie C, Derache A, Aime C, et al. Comparison of two genotypic algorithms to determine HIV-1 tropism. *HIV Med.* 2008; 9:1–5. [PubMed: 18199166]
254. Jensen MA, van 't Wout AB. Predicting HIV-1 coreceptor usage with sequence analysis. *AIDS Rev.* 2003; 5:104–12. [PubMed: 12876899]
255. Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol.* 2006; 80:4698–704. [PubMed: 16641263]
256. Cilliers T, Nhlapo J, Coetzer M, et al. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol.* 2003; 77:4449–56. [PubMed: 12634405]
257. Coetzer M, Cilliers T, Ping LH, Swanstrom R, Morris L. Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates. *Virology.* 2006; 356:95–105. [PubMed: 16942785]
258. Ping LH, Nelson JA, Hoffman IF, et al. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J Virol.* 1999; 73:6271–81. [PubMed: 10400718]
259. Pillai, S.; Good, Benjamin; Richman, Douglas; Corbeil, Jacques. WetCat. 2003. Available from: <http://genomiac2.ucsd.edu:8080/wetcat/v3.html>
260. Poveda E, Seclen E, Gonzalez Mdel M, et al. Design and validation of new genotypic tools for easy and reliable estimation of HIV tropism before using CCR5 antagonists. *J Antimicrob Chemother.* 2009; 63:1006–10. [PubMed: 19261623]
261. Beerenwinkel N, Daumer M, Oette M, et al. Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res.* 2003; 31:3850–5. [PubMed: 12824435]
262. Beerenwinkel N, Schmidt B, Walter H, et al. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. *Proc Natl Acad Sci U S A.* 2002; 99:8271–6. [PubMed: 12060770]
263. Beerenwinkel N, Sing T, Lengauer T, et al. Computational methods for the design of effective therapies against drug resistant HIV strains. *Bioinformatics.* 2005; 21:3943–50. [PubMed: 16144807]
264. Infomatik M-P-I. Geno2Pheno. 2011. [updated December 23, 2011]. Available from: <http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>
265. Lehmann C, Daumer M, Boussaad I, et al. Stable coreceptor usage of HIV in patients with ongoing treatment failure on HAART. *J Clin Virol.* 2006; 37:300–4. [PubMed: 17005445]

266. Domingo, EHJ.; Biebricher, CK.; Eigen, M. Quasispecies: The concept and the word. In: Adrian, J.; Gibbs, CHC.; Garcia-Arenal, Fernando, editors. *Molecular Basis of Virus Evolution*. UK: Cambridge University Press; 1995. p. 171-80.
267. Sander O, Sing T, Sommer I, et al. Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. *PLoS Comput Biol*. 2007; 3:e58. [PubMed: 17397254]
268. Dybowski JN, Heider D, Hoffmann D. Prediction of Co-Receptor Usage of HIV-1 from Genotype. *PLoS Comput Biol*. 2010; 6:e1000743. [PubMed: 20419152]
269. Bozek K, Lengauer T, Sierra S, Kaiser R, Domingues FS. Analysis of physicochemical and structural properties determining HIV-1 coreceptor usage. *PLoS Comput Biol*. 2013; 9:e1002977. [PubMed: 23555214]
270. Canutescu AA, Shelenkov AA, Dunbrack RL Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci*. 2003; 12:2001–14. [PubMed: 12930999]
271. Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol*. 2002; 323:387–406. [PubMed: 12381328]
272. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993; 234:779–815. [PubMed: 8254673]
273. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*. 2001; 98:10037–41. [PubMed: 11517324]
274. Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.
275. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982; 157:105–32. [PubMed: 7108955]
276. Brumme ZL, Dong WW, Yip B, et al. Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. *AIDS*. 2004; 18:F1–9. [PubMed: 15090786]
277. Koot M, van 't Wout AB, Kootstra NA, et al. Relation between changes in cellular load, evolution of viral phenotype, and the clonal composition of virus populations in the course of human immunodeficiency virus type 1 infection. *J Infect Dis*. 1996; 173:349–54. [PubMed: 8568295]
278. Philpott SM. HIV-1 coreceptor usage, transmission, and disease progression. *Curr HIV Res*. 2003; 1:217–27. [PubMed: 15043204]
279. Brumme ZL, Goodrich J, Mayer HB, et al. Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naïve individuals. *J Infect Dis*. 2005; 192:466–74. [PubMed: 15995960]
280. Wilkin TJ, Su Z, Kuritzkes DR, et al. HIV type 1 chemokine coreceptor use among antiretroviral-experienced patients screened for a clinical trial of a CCR5 inhibitor: AIDS Clinical Trial Group A5211. *Clin Infect Dis*. 2007; 44:591–5. [PubMed: 17243065]
281. Traub MA, Icard V, Scholtes C, et al. Discordance in HIV-1 co-receptor use prediction by different genotypic algorithms and phenotype assay: intermediate profile in relation to concordant predictions. *J Med Virol*. 2012; 84:402–13. [PubMed: 22246825]
282. Abbate I, Vlasi C, Rozera G, et al. Detection of quasispecies variants predicted to use CXCR4 by ultra-deep pyrosequencing during early HIV infection. *AIDS*. 2011; 25:611–7. [PubMed: 21160417]
283. Hunt PW, Harrigan PR, Huang W, et al. Prevalence of CXCR4 tropism among antiretroviral-treated HIV-1-infected patients with detectable viremia. *J Infect Dis*. 2006; 194:926–30. [PubMed: 16960780]
284. Chuecal NLM, Alvarez M, Pena A, Guillot V, Garcia-Casas V, Sanchez R, Hernandez-Quero J, Maroto MC, Garcia F. A combination of bioinformatic tools can be accurately used for the screening of coreceptor usage in clinical samples. *Antivir Ther*. 2008; 13:A106.
285. Poveda E, Briz V, Roulet V, et al. Correlation between a phenotypic assay and three bioinformatic tools for determining HIV co-receptor use. *AIDS*. 2007; 21:1487–90. [PubMed: 17589199]

286. Diez-Fuertes F, Delgado E, Vega Y, et al. Improvement of HIV-1 coreceptor tropism prediction by employing selected nucleotide positions of the env gene in a Bayesian network classifier. *J Antimicrob Chemother.* 2013
287. Delgado E, Fernandez-Garcia A, Vega Y, et al. Evaluation of genotypic tropism prediction tests compared with *in vitro* co-receptor usage in HIV-1 primary isolates of diverse subtypes. *J Antimicrob Chemother.* 2012; 67:25–31. [PubMed: 22010208]
288. Fouchier RA, Brouwer M, Broersen SM, Schuitemaker H. Simple determination of human immunodeficiency virus type 1 syncytium-inducing V3 genotype by PCR. *J Clin Microbiol.* 1995; 33:906–11. [PubMed: 7790458]
289. Raymond S, Delobel P, Rogez S, et al. Genotypic prediction of HIV-1 CRF01-AE tropism. *J Clin Microbiol.* 2013; 51:564–70. [PubMed: 23224099]
290. Dorr P, Westby M, Dobbs S, et al. 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.* 2005; 49:4721–32. [PubMed: 16251317]
291. Cabral GB, Ferreira JL, Coelho LP, et al. Concordance of HIV type 1 tropism phenotype to predictions using web-based analysis of V3 sequences: composite algorithms may be needed to properly assess viral tropism. *AIDS Res Hum Retroviruses.* 2012; 28:734–8. [PubMed: 21919801]
292. Sanchez V, Masia M, Robledano C, et al. Performance of genotypic algorithms for predicting HIV-1 tropism measured against the enhanced-sensitivity Trofile coreceptor tropism assay. *J Clin Microbiol.* 2010; 48:4135–9. [PubMed: 20861336]
293. Recordon-Pinson P, Soulie C, Flandre P, et al. Evaluation of the genotypic prediction of HIV-1 coreceptor use versus a phenotypic assay and correlation with the virological response to maraviroc: the ANRS GenoTropism study. *Antimicrob Agents Chemother.* 2010; 54:3335–40. [PubMed: 20530226]
294. Symons J, Vandekerckhove L, Paredes R, et al. Impact of triplicate testing on HIV genotypic tropism prediction in routine clinical practice. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases.* 2012; 18:606–12.
295. Raymond S, Delobel P, Chaix ML, et al. Genotypic prediction of HIV-1 subtype D tropism. *Retrovirology.* 2011; 8:56. [PubMed: 21752271]
296. Wambui V, Kiptoo M, Kinyua J, et al. Predicted HIV-1 coreceptor usage among Kenya patients shows a high tendency for subtype d to be cxcr4 tropic. *AIDS research and therapy.* 2012; 9:22. [PubMed: 22838398]
297. Thielen A, Lengauer T, Swenson LC, et al. Mutations in gp41 are correlated with coreceptor tropism but do not improve prediction methods substantially. *Antivir Ther.* 2011; 16:319–28. [PubMed: 21555814]
298. Hemelaar J, Gouws E, Ghys PD, et al. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS.* 2011; 25:679–89. [PubMed: 21297424]
299. Fernandes AD, Gloor GB. Mutual information is critically dependent on prior assumptions: would the correct estimate of mutual information please identify itself? *Bioinformatics.* 2010; 26:1135–9. [PubMed: 20236946]
300. Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. *Statistical methods in medical research.* 2004; 13:227–38. [PubMed: 15198488]
301. Batorsky R, Kearney MF, Palmer SE, et al. Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proc Natl Acad Sci U S A.* 2011; 108:5661–6. [PubMed: 21436045]
302. Abrahams MR, Treurnicht FK, Ngandu NK, et al. Rapid, complex adaptation of transmitted HIV-1 full-length genomes in subtype C-infected individuals with differing disease progression. *AIDS.* 2013; 27:507–18. [PubMed: 23370465]
303. Henn MR, Boutwell CL, Charlebois P, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* 2012; 8:e1002529. [PubMed: 22412369]

304. Hoffman NG, Seillier-Moiseiwitsch F, Ahn J, Walker JM, Swanstrom R. Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop. *J Virol.* 2002; 76:3852–64. [PubMed: 11907225]
305. Monno L, Saracino A, Scudeller L, et al. Impact of mutations outside the V3 region on coreceptor tropism phenotypically assessed in patients infected with HIV-1 subtype B. *Antimicrob Agents Chemother.* 2011; 55:5078–84. [PubMed: 21876051]
306. Moore JP, Willey RL, Lewis GK, Robinson J, Sodroski J. Immunological evidence for interactions between the first, second, and fifth conserved domains of the gp120 surface glycoprotein of human immunodeficiency virus type 1. *J Virol.* 1994; 68:6836–47. [PubMed: 7933065]
307. Moore JP, Thali M, Jameson BA, et al. Immunochemical analysis of the gp120 surface glycoprotein of human immunodeficiency virus type 1: probing the structure of the C4 and V4 domains and the interaction of the C4 domain with the V3 loop. *J Virol.* 1993; 67:4785–96. [PubMed: 7687303]
308. Wyatt R, Thali M, Tilley S, et al. Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to a component of the CD4 binding site in the fourth conserved region. *J Virol.* 1992; 66:6997–7004. [PubMed: 1279195]
309. Shieh JT, Martin J, Baltuch G, Malim MH, Gonzalez-Scarano F. Determinants of syncytium formation in microglia by human immunodeficiency virus type 1: role of the V1/V2 domains. *J Virol.* 2000; 74:693–701. [PubMed: 10623731]
310. Dimonte S, Mercurio F, Svicher V, et al. Selected amino acid mutations in HIV-1 B subtype gp41 are associated with specific gp120v(3) signatures in the regulation of co-receptor usage. *Retrovirology.* 2011; 8:33. [PubMed: 21569409]
311. Dimonte S, Babakir-Mina M, Mercurio F, et al. Selected amino acid changes in HIV-1 subtype-C gp41 are associated with specific gp120(V3) signatures in the regulation of co-receptor usage. *Virus Res.* 2012; 168:73–83. [PubMed: 22732432]
312. Anastassopoulou CG, Ketas TJ, Klasse PJ, Moore JP. Resistance to CCR5 inhibitors caused by sequence changes in the fusion peptide of HIV-1 gp41. *Proc Natl Acad Sci U S A.* 2009; 106:5318–23. [PubMed: 19289833]
313. Pfaff JM, Wilen CB, Harrison JE, et al. HIV-1 resistance to CCR5 antagonists associated with highly efficient use of CCR5 and altered tropism on primary CD4+ T cells. *J Virol.* 2010; 84:6505–14. [PubMed: 20410277]
314. Garcia JV, Miller AD. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature.* 1991; 350:508–11. [PubMed: 2014052]
315. Amadori A, De Silvestro G, Zamarchi R, et al. CD4 epitope masking by gp120/anti-gp120 antibody complexes. A potential mechanism for CD4+ cell function down-regulation in AIDS patients. *J Immunol.* 1992; 148:2709–16. [PubMed: 1374095]
316. Hubert P, Bismuth G, Korner M, Debre P. HIV-1 glycoprotein gp120 disrupts CD4-p56lck/CD3-T cell receptor interactions and inhibits CD3 signaling. *Eur J Immunol.* 1995; 25:1417–25. [PubMed: 7774645]
317. Iafrate AJ, Bronson S, Skowronski J. Separable functions of Nef disrupt two aspects of T cell receptor machinery: CD4 expression and CD3 signaling. *EMBO J.* 1997; 16:673–84. [PubMed: 9049297]
318. Roshal M, Zhu Y, Planelles V. Apoptosis in AIDS. *Apoptosis: an international journal on programmed cell death.* 2001; 6:103–16. [PubMed: 11321033]
319. Papkalla A, Munch J, Otto C, Kirchhoff F. Nef enhances human immunodeficiency virus type 1 infectivity and replication independently of viral coreceptor tropism. *J Virol.* 2002; 76:8455–9. [PubMed: 12134048]
320. Wilkin TJ, Gulick RM. CCR5 antagonism in HIV infection: current concepts and future opportunities. *Annual review of medicine.* 2012; 63:81–93.
321. Vranken WF, Budesinsky M, Fant F, Boulez K, Borremans FA. The complete Consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character in solution. *FEBS Lett.* 1995; 374:117–21. [PubMed: 7589496]

322. Jmol: an open-source Java viewer for chemical structures in 3D. Available from: <http://www.jmol.org/>
323. PepDraw: a tool to draw peptide primary structure and calculate theoretical properties. Available from: <http://www.tulane.edu/~biochem/WW/PepDraw/>
324. Connell BJ, Michler K, Capovilla A, et al. Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. *AIDS*. 2008; 22:896–9. [PubMed: 18427209]
325. Mulinge M, Lemaire M, Servais JY, et al. HIV-1 Tropism Determination Using a Phenotypic Env Recombinant Viral Assay Highlights Overestimation of CXCR4-Usage by Genotypic Prediction Algorithms for CRRF01_AE and CRF02_AG. *PLoS One*. 2013; 8:e60566. [PubMed: 23667426]
326. Garrido C, Roulet V, Chueca N, et al. Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. *J Clin Microbiol*. 2008; 46:887–91. [PubMed: 18199789]
327. Chueca N, Garrido C, Alvarez M, et al. Improvement in the determination of HIV-1 tropism using the V3 gene sequence and a combination of bioinformatic tools. *J Med Virol*. 2009; 81:763–7. [PubMed: 19319937]

Fig. 1. HIV-1 entry mechanism

HIV-1 entry has been shown to initially require the binding of trimeric gp120 to the host protein CD4 on the target cell plasma membrane. This interaction triggers a conformational change in the HIV-1 envelope protein that results in the enhanced exposure of the gp120 V3 loop, which was initially concealed by the V1/V2 region. The V3 loop subsequently engages one of two chemokine co-receptors, CXCR4 (left) or CCR5 (right). The overall charge of the V3 loop largely determines co-receptor usage, with CXCR4-utilizing virus having a greater net positive charge. In addition, in CXCR4-utilizing viruses the V3 loop contacts primarily the N-terminus and ECL-2 of CXCR4, while CCR5-utilizing viruses contact the N-terminus and ECL-1 of CCR5. Finally, co-receptor binding initiates the membrane fusion machinery of HIV-1 gp41.

Fig. 2. Structural and chemical composition of the V3 loop

The secondary structure and Van der Waals surface of the gp120 V3 loop (PDB ID 1CE4) is displayed using Jmol (left) [321, 322]. Helical structures are indicated in purple and pink and unstructured regions of the V3 loop are displayed in white. Representative peptide primary structures were generated using the PepDraw online tool (right) [323]. For CXCR4-utilizing V3, the HXB2 strain was used while for the CCR5-utilizing V3 the BaL strain was used. Charged side chains are indicated using boxes, with yellow indicating positively charged amino acids (R and K) and green indicating negatively charged amino acids (D and E). The white boxes represent any amino acid residue. The position of these residues is provided based on the alignment of the sequences against consensus subtype B sequence from LANL. The location of these residues is also highlighted in the three-dimensional structure of the V3 loop. In addition, positions 11 and 25 are indicated in a box due to their prominence in *in silico* prediction of co-receptor utilization.

Fig. 3. Overview of HIV-1 envelope co-receptor utilization evolution in periphery and CNS

A. Transmitted HIV-1 quasiespecies infect CD4⁺ T cells and cells of the monocyte-macrophage lineage within the periphery (1). Following initial infection, a genetic bottleneck results in a clonal population predominantly exhibiting CCR5 co-receptor utilization (2). The founder swarm traffics to the CNS early in infection *via* infected monocytes (3) and can also traverse the blood-brain barrier as free virus (3). The brain founder swarm (4) is considered to be R5-utilizing and M-tropic on the basis of the cellular milieu within the CNS. Within the CNS, microglia and perivascular macrophages predominantly produce new virus while astrocytes are also infected, however this infection results in very limited virus production (5). Throughout chronic infection, virus may continue to traffic between physiological compartments (7). The trafficking between the periphery and CNS likely increases in magnitude over time (8) due to the leakiness of the blood-brain barrier during advanced neurological disease. Within the periphery, CD4⁺ T cells remain the predominant infected cell type and driver of viral evolution. As HIV-1 disease progresses, the emergence of X4- and X4/R5-using quasiespecies is commonly observed (9). In late infection, many patients are observed to experience a tropism switch to a predominantly X4-utilizing swarm (10), however this switch is not universally observed and many patients remain predominantly R5-utilizing (11).

B. Macrophage tropism is considered to occupy a spectrum of HIV-1 envelope genotypes and phenotypes that are predominantly CCR5-utilizing but also include dual-tropic and CXCR4-utilizing species. This differs from T-cell tropism in that most isolates, regardless of co-receptor usage, should be able to efficiently infect T cells. The proportion of T cells infected may vary based on co-receptor expression, however a robust infection should still develop in a majority of infected individuals.

Fig. 4. Identification of amino acid positions that co-selected with V3 in subtype B and C Arcs between V3 positions and gp120, gp41, Nef, and the LTR represent co-evolving residues as defined by statistically significant mutual information (MI). Each row represents the subset of LANL sequences that were subtype B, subtype C, or either subtype B or C. Each column represents the subset of sequences that are CCR5 utilizing (R5), CXCR4 utilizing (X4), or either X4 or R5 utilizing as predicted by WebPSSM. The combined column shows the superimposed links with R5 in red, X4 in blue, and all-patients in black. Subsets that had fewer than 10 X4 or 10 R5 sequences of co-linear sequence were excluded from the analysis and the region was removed from the diagram.

Fig. 5. Differential amino acid distribution (Circos diagram)

Arcs between V3 positions and gp120, gp41, Nef, and the LTR that have differential residue preference as calculated by a Fisher's Exact test. Each Circos diagram represents a subset of LANL sequences that were subtype B alone, subtype C alone, or either combination of subtype B and C. The combined figure shows the superimposed links with subtype B in green, subtype C in red, and subtype B and C in blue. Subsets that had fewer than 10 X4 or 10 R5 sequences of co-linear sequence were excluded from the analysis and the region was removed from the diagram.

Table 1Evaluation of Web-based *in silico* co-receptor usage prediction tools.

<i>In Silico</i> Online Tool	WetCat^a	WebPSSM^b	Geno2pheno^c
Method Utilized	<ul style="list-style-type: none"> SVM (support vector machines) Charge rule Quinlan C4.5 PART (partial decision trees) 	<ul style="list-style-type: none"> PSSM (position-specific scoring matrices) 	<ul style="list-style-type: none"> SVM (support vector machine)
Training Sequences	<ul style="list-style-type: none"> V3 sequences containing 34–36 amino acid residues from LANL HIV Sequence Database 168 R5, 103 X4, and 21 dual tropic 	<ul style="list-style-type: none"> 70 SI and 187 NSI sequences from 107 subjects 168 R5, 17 X4, and 28 dual-tropic sequences from 177 subjects 	<ul style="list-style-type: none"> 1,100 Clonal sequences (769 R5, 210 X4, and 131 dual-tropic) from 332 patients 920 naïve therapy patients–derived samples
Information Input	<ul style="list-style-type: none"> V3 sequence 	<ul style="list-style-type: none"> Up to 1,000 V3 FASTA formatted sequences per upload 35 amino acids in length; if not, multiple output results will be reported X4R5 or SI/NSI matrix choices Subtype B or subtype C matrix choices 	<ul style="list-style-type: none"> V3 sequence in either plain text or FASTA format File can be uploaded Clinical parameter can be optional Conservative level can be adjusted Choice of prediction method
Information Output	<ul style="list-style-type: none"> Classifies sequences as X4 or non-X4 	<ul style="list-style-type: none"> Provides a quantitative score indicating the likelihood of CXCR4 utilization Percentile 11/25 rule genotype Positive charge Negative charge 	<ul style="list-style-type: none"> Input V3 sequence alignment to HXB2 Prediction whether CXCR4 can be used Possibility of CCR5 inhibitor administration V3 subtype identification
Advantages	<ul style="list-style-type: none"> User can choose any type of analysis 	<ul style="list-style-type: none"> PSSM score predicts viral evolutionary phenotype, applying in tracking therapy response PSSM-B offers analysis for both co-receptor (CXCR4 <i>vs</i> CCR5) and syncytia induction ability (SI <i>vs</i> NSI) Viral tropism of HIV-1 subtype C can be determined PSSM-C algorithm exhibits 100% accuracy [324] 	<ul style="list-style-type: none"> Input sequence can be any kind of format Offers options to incorporate any clinical data for prediction which increase 10% sensitivity of prediction [232, 234] Two trained data set choice of use (clonal <i>vs</i> clinical) Provide HIV-1 subtype identification result

<i>In Silico</i> Online Tool	WetCat^a	WebPSSM^b	Geno2pheno^c
		<ul style="list-style-type: none"> High accuracy (80%) in non-subtype B co-receptor prediction [325] 	
Concordance with Trofile TM [260]	<ul style="list-style-type: none"> 68% (SVM) 74% (Charge rule) 76% (C4.5) 75% (PART) 	<ul style="list-style-type: none"> 80% (X4/R5 matrix) 79% (SI/NSI matrix) 	<ul style="list-style-type: none"> 75% (1% False positive rate, FPR) 76% (5% FPR) 71% (10% FPR) 67% (15% FPR) 66% (20% FPR)
Concordance of non-subtype B prediction, with phenotypic assay [326]	<ul style="list-style-type: none"> 78.8% (SVM) 82.5% (Charge rule) 77.5% (C4.5) 82.5% (PART) 	<ul style="list-style-type: none"> 82.5% (X4/R5 matrix) 83.2% (SI/NSI matrix) 	<ul style="list-style-type: none"> 71.3%
Pitfalls	<ul style="list-style-type: none"> Lower sensitivity [234] Mix subtypes of V3 sequences within training data [248] Website is not stable 	<ul style="list-style-type: none"> Input V3 sequence which is not 35 amino acid residues size yields multiple scores and predictions Input sequences has to be only in FASTA format 	<ul style="list-style-type: none"> Unspecified subtype of V3 sequence within training data set Higher discordance with phenotypic assay in determination of X4 user within subtype C as compared to <i>vs</i> PSSM-C_{SI/NSI} (22.5% 6.3% discordance) [326]
Reference	[234, 248]	[251, 255, 324, 325]	[231, 232, 234, 326]

^a <http://genomiac2.ucsd.edu:8080/wet-cat/v3.html>

^b <http://indra.mullins.microbiol.washington.edu/webpssm/>

^c <http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>

Table 2Web-based *in silico* X4/R5 prediction performance in different HIV-1 subtypes.

Subtype/CRF	Wet-Cat ^a	WebPSSM ^b	Geno2pheno ^c
A	N/A *	100% sensitivity [287] 100% specificity [287]	100% sensitivity [287] 93% specificity [287]
CRF01_AE	N/A *	N/A *	91% sensitivity [289] 54% specificity [289]
B	98.8% sensitivity [327] 62.5% specificity [327]	82.5% sensitivity [327] 97.5% specificity [327]	91.2% sensitivity [327] 86.6% specificity [327]
B	N/A *	100% sensitivity [287] 100% specificity [287]	100% sensitivity [287] 90% specificity [287]
B	67% sensitivity [254] 93% specificity [254]	62% sensitivity [254] 95% specificity [254]	N/A *
B	21.8% sensitivity [234] 89.6% specificity [234]	24.5% sensitivity [234] 96.9% specificity [234]	44.7% sensitivity [234] 90.6% specificity [234]
B	100% sensitivity [326] 80% specificity [326]	90% sensitivity [326] 92% specificity [326]	90% sensitivity [326] 88% specificity [326]
C	N/A *	83.3% sensitivity ** [255] 83% specificity ** [255]	N/A *
D	N/A *	N/A *	75% sensitivity [295] 54% specificity [295]
F	N/A *	50% sensitivity [287] 100% specificity [287]	75% sensitivity [287] 90% specificity [287]
G	N/A *	100% sensitivity [287] 100% specificity [287]	100% sensitivity [287] 90% specificity [287]

^aSVM classifier is chosen to represent wet-cat performance.^bX4/R5 matrix, otherwise as indicated.^c10% FPR, clonal data set, otherwise as indicated.

* This method was not used in analysis of indicated subtype.

** Verified by SI/NSI matrix from PSSM-C.

Table 3

Differential amino acid analysis comparing between the X4 and R5 viruses.

Location	Position	p-Value		
		Subtype B	Subtype C	Subtype B and C
LTR	239	0.00497	N/A *	0.03226
Nef	18	0.00814	0.14454	0.14759
Nef	19	0.00318	0.33431	0.01219
Nef	43	0.00855	1.00000	0.05604
Nef	176	0.59968	0.00481	0.12576
Nef	18	0.00814	0.14454	0.14759
gp120	2	0.00115	0.02672	0.00005
gp120	39	1.00000	0.00309	0.01414
gp120	46	0.05015	0.21926	0.00771
gp120	48	1.00000	0.00930	0.02629
gp120	70	0.01047	0.08578	0.00174
gp120	152	0.44960	0.00813	0.06693
gp120	158	0.00456	0.75925	0.00104
gp120	190	0.00226	1.00000	0.02002
gp120	251	0.00001	0.08578	4.71428E-06
gp120	257	3.99915E-08	0.29974	2.02343E-07
gp120	259	0.00990	0.59986	0.02009
gp120	265	0.00226	0.07233	0.00056
gp120	273	1.00000	0.00441	0.05694
gp120	276	0.72339	0.00295	0.09551
gp120	285	0.00178	0.62499	0.00025
gp120	289	0.05053	0.15592	0.00511
gp120	297	0.00131	0.59379	0.00347
gp120	302	0.06110	0.12714	0.00931
gp120	341	0.45241	0.00694	0.03648

Location	Position	p-Value		
		Subtype B	Subtype C	Subtype B and C
gp120	343	0.02230	0.03656	0.00214
gp120	389	0.00688	1.00000	0.02025
gp120	412	0.00998	1.00000	0.03176
gp120	426	0.08659	0.00819	0.00237
gp120	432	0.06961	0.00798	0.00413
gp120	440	0.00419	0.61557	0.02745
gp41	47	0.00011	0.16714	0.15052
gp41	162	0.89417	0.00614	0.16451
gp41	177	0.20168	0.00786	0.00379
gp41	273	1.00000	0.00607	0.02399
gp41	291	0.51588	0.00151	0.88597
gp41	303	0.11419	0.00652	1.00000
gp41	307	0.03615	1.00000	0.00799

The frequency of change at indicated amino acid position within each genes/promoter was calculated in reference to the conB sequence. Fisher Exact test was then utilized to compare the difference between the X4 and the R5 viruses. p<0.05 is considered significant.

* Insufficient data for analysis.