Edinburgh Research Explorer

Intercompartmental Recombination of HIV-1 Contributes to env Intrahost Diversity and Modulates Viral Tropism and Sensitivity to Entry Inhibitors

Citation for published version:

Digital Object Identifier (DOI):
10.1128/JVI.00131-11

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
Journal of Virology

Publisher Rights Statement:
RoMEO blue

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Intercompartmental Recombination of HIV-1 Contributes to env Intrahost Diversity and Modulates Viral Tropism and Sensitivity to Entry Inhibitors

Richard J. P. Brown,† Paul J. Peters,‡ Catherine Caron, Maria Paz Gonzalez-Perez, Leanne Stones, Chiambah Ankghuambom, Kemebradikumo Pondei, C. Patrick McClure, George Alemnji, Stephen Taylor, Paul M. Sharp, Paul R. Clapham, and Jonathan K. Ball

School of Molecular Medical Sciences, The University of Nottingham, Queen’s Medical Centre, Nottingham NG7 2UH, United Kingdom; Center for AIDS Research, Program in Molecular Medicine and Department of Molecular Genetics and Microbiology, 373 Plantation Street, University of Massachusetts Medical School, Worcester, Massachusetts 01605; Faculty of Medicine and Bio-Medical Sciences, University of Yaounde I, Yaounde, Cameroon; Department of Sexual Health and HIV Medicine, Directorate of Infection, Birmingham Heartlands Hospital, Heart of England NHS Foundation Trust, Birmingham B9 5SS, United Kingdom; and Institute of Evolutionary Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JT, United Kingdom

Received 19 January 2011/Accepted 30 March 2011

HIV-1 circulates within an infected host as a genetically heterogeneous viral population. Viral intrahost diversity is shaped by substitutional evolution and recombination. Although many studies have speculated that recombination could have a significant impact on viral phenotype, this has never been definitively demonstrated. We report here phylogenetic and subsequent phenotypic analyses of envelope genes obtained from HIV-1 populations present in different anatomical compartments. Assessment of env compartmentalization from immunologically discrete tissues was assessed utilizing a single genome amplification approach, minimizing in vitro-generated artifacts. Genetic compartmentalization of variants was frequently observed. In addition, multiple incidences of intercompartment recombination, presumably facilitated by low-level migration of virus or infected cells between different anatomic sites and coinfection of susceptible cells by genetically divergent strains, were identified. These analyses demonstrate that intercompartment recombination is a fundamental evolutionary mechanism that helps to shape HIV-1 env intrahost diversity in natural infection. Analysis of the phenotypic consequences of these recombination events showed that genetic compartmentalization often correlates with phenotypic compartmentalization and that intercompartment recombination results in phenotype modulation. This represents definitive proof that recombination can generate novel combinations of phenotypic traits which differ subtly from those of parental strains, an important phenomenon that may have an impact on antiviral therapy and contribute to HIV-1 persistence in vivo.

Approximately 33 million people globally are infected with HIV-1 group M viruses, with an estimated 2.5 million new infections in 2007 (87). The extreme diversity of HIV-1 is characterized by nine phylogenetically distinct subtypes and a multitude of circulating recombinant forms (CRFs) (6). Such diversity poses a significant obstacle for the development of a successful prophylactic vaccine (21, 82). HIV-1 circulates within an infected host as a genetically heterogeneous viral population. The rapid rate of HIV-1 diversification within an infected individual is driven, in part, by the error-prone nature of HIV-1 reverse transcription and the high replicative rate in vivo that facilitates the incremental accumulation of insertions, deletions (indels), and point mutations in the viral genome (36). The inherent genetic variability of the population leads to humoral (86) and cellular (31) immune escape, resistance to antiretrovirals (9), and altered cytopathogenicity and cell tropism (48–50). Compartmentalization of HIV-1 populations derived from a range of tissues has been reported, including blood versus brain/cetal nervous system (1, 27, 47, 53, 67), brain versus lymph (4, 42, 68), blood versus female genital tract (51), and blood versus semen (10, 12, 24, 52, 93). In addition, genetic segregation of HIV-1 variants is also apparent between different populations of Langerhans cells (61) and at different sites within the brain (63), spleen (11), and gut (79). Such compartmentalized variation is shaped by a combination of founder effect, restricted cellular/viral trafficking between segregated subpopulations and local environmental selective pressures, including receptor/cellular tropism and tissue-specific immune responses.

A large part of HIV-1’s genomic variability is located within the env gene, which encodes the transmembrane glycoprotein gp41 and surface glycoprotein gp120. These glycoproteins exist as trimeric spikes of gp41-gp120 heterodimers on the surface of the viral membrane (34, 92). Progressive evolution of the env gene contributes to escape from host immune responses. However, these processes are restricted by functional constraints associated with receptor and coreceptor binding and membrane fusion. The interaction of gp120 with the primary
cellular receptor CD4 results in major conformational changes in the env trimer (34), which facilitates subsequent binding to the chemokine coreceptors CCR5 (8) and/or CXCR4, followed by gp41-mediated membrane fusion.

HIV-1 genetic heterogeneity is also shaped by recombination and mosaicism. HIV-1 is a highly recombinogenic virus, with exchange of genetic material between divergent group M subtypes giving rise to multiple CRFs worldwide (6, 56, 57). Examples of both inter- and intrasubtype recombinants have been reported (2, 6, 35, 58). Molecular epidemiological studies demonstrate rapid dissemination of certain CRFs in specific geographical localities (33, 77), suggesting that in some settings a fitness advantage may be conferred by mosaic genomes (3). However, this has not been definitively demonstrated. Recombination occurs via reverse transcriptase template switching during proviral DNA synthesis. Consequently, this phenomenon can only occur when a target cell is dually infected by genetically discrete viral strains (36, 56, 57). At the intrapatient level, recombination events continually produce mosaic genomes (7, 45). The rate of HIV-1 recombination has been shown to significantly exceed the rate of nucleotide misincorporation by the viral reverse transcriptase per round of genome replication (25, 39). Moreover, rates of HIV-1 recombination have been shown to differ between infected cell types (32).

Intrapatient recombination has been proposed to facilitate recombination by the viral reverse transcriptase per round of genome replication (25, 39). Moreover, rates of HIV-1 recombination have been shown to differ between infected cell types (32).

The presence of intercompartment recombinant viruses. Importantly, these recombination events conferred altered and potentially advantageous phenotypes compared to compartmentalized nonrecombinant strains.

### MATERIALS AND METHODS

**Sample collection and preparation.** Paired blood and semen samples were acquired from four treatment-naïve HIV-1+ men attending clinics in Tiko, Cameroon, and Birmingham, United Kingdom. Blood and seminal plasmas were prepared by centrifugation at 1,500 \( \times g \) for 20 min and stored at \(-70^\circ\)C prior to RNA extraction. Seminal fluid mononuclear cells (SFMC) and peripheral blood mononuclear cells (PBMC) were isolated by centrifugation and Ficoll density gradient centrifugation, respectively. Washed PBMC and SFMC pellets were resuspended in 250 \( \mu L \) of saline prior to total DNA extraction (22). Paired samples were collected on the same day for each individual patient and processed within 4 h. Brain and cervical lymph node tissue were obtained at autopsy in Birmingham, United Kingdom, from a single HIV-1+ patient who died of AIDS related complications and stored at \(-70^\circ\)C prior to total DNA extraction. Sample collection and subsequent research on all patient tissue utilized in the present study received ethical approval by the West Midlands Local Ethics Committee (United Kingdom) and the National Ethics Committee (Cameroon). All participants gave written informed consent.

**Nucleic acid extraction and cDNA synthesis.** Viral RNA was extracted from blood and seminal plasma samples by using a QiAamp viral RNA kit (Qiagen) according to the manufacturer’s instructions. Total DNA was extracted from PBMC, SFMC, brain autopsy, and lymph node autopsy samples by using a Stratagene DNA extraction kit according to a previously described, modified version of the manufacturer’s protocol (4, 41). A gene-specific primer (20) was used to reverse transcribe cDNA from each RNA sample using a Thermoscript kit (Invitrogen) according to the manufacturer’s instructions.

**PCR amplification of env from single molecule templates.** Due to high levels of circulating HIV-1 diversity apparent in Cameroon, nested primer pairs shown to be effective at amplifying env from a diverse range of subtypes were used (20) according to previously described conditions (62). Accurate and representative sampling of the env intrahost diversity was achieved via SGA to preclude generation of in vitro artifacts (42, 62, 70). Amplicons were sequenced via BigDye terminator chemistry using an ABI 3100 capillary sequencer. Sequence chromatographs were manually inspected, and amplicons generated from \( >1 \) DNA/provirus molecule were excluded from subsequent analyses. Amplicons that showed evidence for APOBEC3G-induced G-to-A hypermutation (37, 59, 65) or exhibited \( >1 \) example of a translation terminating stop codon or a frameshift inducing indel were also omitted. Subtype/CRF designations were determined by using the REGA subtyping tool (13). Sequences generated for the present study have been submitted to GenBank under accession numbers JF706370 to JF706501.

**Sequence alignment and phylogenetic reconstruction.** Nucleotide sequences were aligned, with manual editing, according to overlying amino acid sequence using the CLUSTAL W algorithm implemented in MEGA4 software (76). Regions of ambiguous alignment were gap stripped from patient-specific and community diversity.
bined alignments. Pairwise distance matrices were calculated for compartment-specific env alignments via the maximum-composite-likelihood method implemented in MEGA4 software. Values were plotted in Prism GraphPad. A maximum-likelihood (ML) phylogenetic tree was generated for the combined alignment using the PHYML program (23) under the best-fit model of nucleotide substitution calculated by ModelTest (55) and midpoint rooted. Bootstrap values were derived from 1,000 replicate neighbor-joining trees (60), and significant (≥70%) values were assigned to internal ML tree branches.

Recombinant identification. Patient-specific alignments were subsequently divided into three segments containing equal numbers of variable sites, and 1,000 partition-homogeneity test replicates were conducted using PAUP* v4.0b10 (75). A P value for the congruence of phylogenetic trees derived from different regions of the env gene was subsequently obtained for each patient’s viral population (19). ML trees were then generated for each patient’s three env region alignments under the best-fit model of nucleotide substitution using PAUP* v4.0b10. Topologies were inspected for significant positional switching of isolates between clades in different regions of env. Amplicons that exhibited shifting between compartment-specific clades were further investigated for mosaicism using a combination of pairwise diversity plots and informative site arrays (56, 72) as implemented in SimPlot version 3.5.1 (35). Breakpoint P values were calculated via 1,000,000 random shuffles of the order of sites of each informative sites array, with the significance test asking how often pseudoreplicate arrays are generated with a breakpoint as high as the observed real data. Absolute breakpoint coordinates were derived from diversity plot crossover points over all mosaics composed of identifiable parental strains were characterized.

Cloning of env amplicons for phenotyping. The phenotypic consequences of intercompartment recombination were subsequently investigated in patients E21 and KM34. Although intercompartment mosaicism was detected in patient KM11, we were unable to clone these env amplicons, despite numerous attempts using a variety of cloning procedures. Therefore, we were unable to perform phenotypic characterization of envelope genes from this patient. Previous phenotypic studies of subtype B primary isolate env genes (86) report successful pseudovirion production using env/rev cassettes ligated into appropriate expression vectors. Consequently, env amplicons were fused with HXB2 env1 using an in-house recombinant PCR methodology. Briefly, amplification reactions were set up in 25-μl volumes containing 5 pmol of primer HXB2-R_E_F (5′-CACACACTACCTTTTGGACCTACCTG-3′), 5 pmol of primer HXB2-R_E_A (5′-TCAACACTAATTGTTGACCTACCTG-3′), 200 mM deoxynucleoside triphosphates, 0.5 U of Phusion high-fidelity DNA polymerase (Finnzymes), 1× Phusion HF buffer, and 250 pg of HXB2 full-length genome containing plasmid template. The PCR cycling parameters consisted of an initial denaturation step at 98°C for 30 s, followed by 35 cycles of 98°C for 10 s, 66°C for 15 s, and 72°C for 15 s, with a final extension step at 72°C for 1 min. Equimolar amounts of HXB2 env and SGA-derived env amplicons were then used as a template for a recombinant PCR amplification using the primers HXB2-R_E_F and env_M (20) according to the reaction setup conditions outlined above. Recombinant PCR was achieved via a two-step cycling strategy. Cycling parameters were an initial denaturation step at 98°C for 30 s, followed by 35 cycles of 98°C for 10 s and 72°C for 45 s, with a final extension step at 72°C for 1 min. The resulting env/rev cassettes were ligated into pcDNA3.1 D-TOPO (Invitrogen), confirmed by sequencing, and transformed in TOP10 or STBL3 cells (Invitrogen).

Cell cultures. HeLa TZM-BL cells were used to titrate env− pseudovirions and to evaluate HIV-1 neutralization (54, 85). Macrophage cultures were prepared from elutriated (26) or blood monocytes (48) as described previously. HEK 293T cells (15) were used to prepare env+ pseudovirions by transfection.

Production and titration of env+ pseudovirions. Both a pNL4.3 construct lacking env and an env+ pcDNA3.1 D-TOPO expression vector were used to produce env+ pseudovirions as described previously (48, 50).

Infectivity assays. Infectivity assays were performed as previously described (16). Primary macrophages were treated with DEAE dextran (10 μg/ml) prior to infection, before addition of an equal volume of serially diluted env− pseudovirions and spinoculation (46). HeLa TZM-BL cells were infected without DEAE dextran or spinoculation.

Inhibition and neutralization assays. Inhibition and neutralization assays for sCD4 (49) and neutralizing monoclonal antibodies b12 (91) and 2G12 (64) were carried out as described previously using HeLa TZM-BL cells as target cells (16, 49). For maraviroc, a CCR5 receptor antagonist (14), cells were treated with 2-fold dilutions in 50 μl for 30 min before adding an equal volume containing 200 FFU of pseudovirus. After 3 h at 37°C, the virus–antibody mixture was removed, growth medium was added, and infected cells were incubated at 37°C for a total of 48 h. To evaluate residual infectivity, the medium was removed, and 100 μl of medium without phenol red was added. The cells were then fixed and solubilized by adding 100 μl of Beta-Glo (Promega, Inc.). Luminescence was then read in a BioTek Clarity luminometer.

RESULTS

Tissue panel and generated env amplicons. Paired HIV-1+ patient tissue samples included blood and semen from treatment-naïve Cameroonian and United Kingdom patients at distinct disease stages, in addition to archival brain and lymph node autopsies samples from a United Kingdom patient who died of AIDS related complications (Table 1). PCR amplification of env genes from single cDNA/proviral template molecules was then performed. A total of 132 env single molecule amplicons were generated, with a range of 20 to 35 amplicons per patient. Multiple sequences from each patient were used to elucidate the infecting subtype/CRF. United Kingdom patients UKBH1 and E21 were infected with subtype B viruses, Cameroonian patients KM11 and KM18 harbored CRF02_AG viruses, while Cameroonian patient KM34 was infected with a group M virus which did not correspond to any described subtype/CRF (Table 1). To ensure patient KM34 was infected with a group M strain, rather than group N or O viruses which also cocirculate in Cameroon, a confirmatory phylogenetic analysis with a panel of HIV-1 group M, N, and O reference strain env genes was conducted. This analysis confirmed KM34 env genes clustered firmly within the group M radiation (data not shown).

Length polymorphism and nucleotide sequence divergence. The variable regions of env exhibit marked length variability and the length of the V1/V2 region has been shown to influence receptor affinity, cellular tropism, and sensitivity to neutralization. Consequently, the median lengths of gp160 proteins encoded by env genes derived from different tissue compartments were compared for each patient’s viral population (see Fig. S1A in the supplemental material). Analysis of compartment-specific viral populations from patient E21 showed that the brain-derived population had more compact env coding regions compared to viruses derived from lymph node. In patient KM18 the median length of semen-derived gp160 was greater than blood-derived gp160, while this trend was reversed for patient KM34. The overall length of the env genes for patients UKBH1 and KM11 were not significantly different between anatomic sites.

Patient-specific alignments were then gap stripped, and pair-

FIG. 1. Phylogenetic assessment of env compartmentalization from immunologically discrete tissues. A midpoint-rooted ML phylogenetic reconstruction for the combined HIV-1 env alignment generated using the GTR+I+Γ substitution model (74, 88, 89) is shown. Branch lengths are in accordance with the scale bar and are proportional to genetic distance. Significance of clustering of isolates was assessed via bootstrap resampling of the sequence data derived from 1000 replications. *, Values >70%; **, values >99%. Isolates marked with a “+” were selected for subsequent phenotyping.
wise nucleotide sequence divergences were calculated for each patient’s compartment-specific pool of env variants. The average substitutions per site were calculated for all sequence pairs derived from each tissue compartment and plotted (see Fig. S1B in the supplemental material). Median pairwise distances were generally lower in brain compared to lymph tissue and in semen compared to blood.

**Phylogenetic assessment of intrahost compartmentalization.** To assess the specific patterns of HIV-1 compartmentalization in each infected individual, phylogenetic analyses were conducted. All env sequences were aligned in a combined data set to enable identification of possible cross-contamination events. The env sequences clustered into five patient-specific monophyletic clades which exhibit various degrees of intrahost compartmentalization (Fig. 1). Patient KM11’s sequences exhibited monophyletic blood- and semen-derived HIV-1 populations and a third population containing sequences derived from both compartments. Topologically, the extent of compartmentalization observed in patient’s E21 and KM34 was identical, with two distinct sequence populations observed. In both patients, all sequences of one type form a clade within the radiation of sequences of the other type. These distinct sequence populations were largely correlated with tissue compartment. KM18 exhibited almost complete compartmentalization of env variants, although low-level migration of virus and/or infected cells between tissue compartments was also detected. Finally, there was no evidence of compartmentalization of blood- and semen-derived sequences obtained from patient UKBH1.

**Intercompartment recombination.** The results of partition incongruence tests (19) showed that different regions of the env gene were topologically incongruent in all patients ($P < 0.05$), which is indicative of recombination events. Therefore, to assess this further, the trees generated during the incongruence analysis were manually inspected for significant positional shifting of isolates between compartment specific clades in different regions of env: a signature of intercompartment recombination. Utilizing this method, we identified a total of eight putative intercompartment recombinants in 3 patients (E21, $n = 4$; KM11, $n = 1$; KM34, $n = 3$; see Fig. S2 in the supplemental material). There was no compelling evidence for intercompartment recombination in patients UKBH1 and KM18. Assignment of parental sequences was achieved via iterative manual generation of informative site arrays (55, 72), with every possible combination of parental sequences tested against each putative mosaic. The two sequences that in combination gave the highest number of informative sites were identical, with two distinct sequence populations observed. In one derived from the lymph tissue of E21, which was representative of lymph sequences that possessed an unusual V3 loop apical sequence with a triplet amino acid insert (GPG), plus a blood-derived sequence from patient KM34, both utilized CXCR4.

**Macrophage tropism.** We next investigated how efficiently R5 gp160s could mediate infectivity of primary macrophages. Figure 4 presents macrophage infectivity of pseudovirions as a percentage of infectivity in HeLa TZM-BL cells. A macrophage infectivity of >1% was considered macrophage-tropic. Compartmentalization of macrophage tropism was evident between lymph node/brain- and blood/semen-derived viruses from patients E21 and KM34, respectively. All but one of the E21 brain-derived gp160s conferred higher macrophage infectivity than those derived from lymph node. The exception was E21BrD9, which exhibited substantially reduced macrophage tropism (0.49%) compared to other brain isolates (7.55 to 25.94%). Comparison of its sequence with the other brain-derived sequences showed that the former had an additional two amino acid deletion and loss of the potential N-linked glycosylation (PNG) site at position N186 in the V2 loop, a region which has been previously shown to be important in conferring macrophage-tropism (83). The two recombinant viruses obtained from E21 brain were both macrophage-tropic. R5 macrophage tropism is modulated by polymorphisms in the CD4 contact residues (30, 49) and residues that affect CD4 binding site exposure. These include residues flanking the CD4 binding site (16), an E153G substitution in V1 (43), and the presence of an asparagine at residue 283 (18, 50). Two CD4 contact residues (residues 281 and 455) (30), plus an additional residue (residue 291) thought to modulate macrophage tropism (16), exhibit compartment specific differences between brain- and lymph-derived sequences (Fig. 5A). Recombinant brain-derived gp160 E21BrD107 is highly macrophage-tropic and yet possesses identical residues to those found in the non-macrophage-tropic lymph-derived sequences at these positions. However, large differences in the length and number of corresponding to each of these fragments can be seen in Fig. 3, with the positional shifting of env E21BrD107 between compartment specific clades highlighted. Patient KM11 harbored a recombinant virus present in blood which possessed a V1/V2 loop (plus flanking regions) homologous to sequences observed in the seminal population. Finally, three further recombinants were identified in patient KM34, where semen variants contained either V1/V2 (KM34SeR63) or C3 (KM34SeR3 and KM34SeRS4) sequences corresponding to blood-derived virus (see Fig. 2).

**Phenotypic characterization.** We next assessed whether there were differences in the phenotypes of viruses present in each tissue compartment and whether the characterized recombination events affected phenotype. Full-length env sequences representative of the tissue compartment specific gp160s, plus the identified intercompartment recombinants from patients E21 and KM34, were cloned into a mammalian expression vector. Resulting gp160 clones were used in fusion and pseudoparticle entry assays to investigate coreceptor usage, macrophage tropism, and sensitivity to a range of entry inhibitors and neutralizing antibodies that block viral entry. A summary of these data can be seen in Table 3. The majority of the clones analyzed used the CCR5 coreceptor. Two clones, one derived from the lymph tissue of E21, which was representative of a minor group of lymph sequences that possessed an unusual V3 loop apical sequence with a triplet amino acid insert (GPG), plus a blood-derived sequence from patient KM34, both utilized CXCR4.

**Macrophage tropism.** We next investigated how efficiently R5 gp160s could mediate infectivity of primary macrophages. Figure 4 presents macrophage infectivity of pseudovirions as a percentage of infectivity in HeLa TZM-BL cells. A macrophage infectivity of >1% was considered macrophage-tropic. Compartmentalization of macrophage tropism was evident between lymph node/brain- and blood/semen-derived viruses from patients E21 and KM34, respectively. All but one of the E21 brain-derived gp160s conferred higher macrophage infectivity than those derived from lymph node. The exception was E21BrD9, which exhibited substantially reduced macrophage tropism (0.49%) compared to other brain isolates (7.55 to 25.94%). Comparison of its sequence with the other brain-derived sequences showed that the former had an additional two amino acid deletion and loss of the potential N-linked glycosylation (PNG) site at position N186 in the V2 loop, a region which has been previously shown to be important in conferring macrophage-tropism (83). The two recombinant viruses obtained from E21 brain were both macrophage-tropic. R5 macrophage tropism is modulated by polymorphisms in the CD4 contact residues (30, 49) and residues that affect CD4 binding site exposure. These include residues flanking the CD4 binding site (16), an E153G substitution in V1 (43), and the presence of an asparagine at residue 283 (18, 50). Two CD4 contact residues (residues 281 and 455) (30), plus an additional residue (residue 291) thought to modulate macrophage tropism (16), exhibit compartment specific differences between brain- and lymph-derived sequences (Fig. 5A). Recombinant brain-derived gp160 E21BrD107 is highly macrophage-tropic and yet possesses identical residues to those found in the non-macrophage-tropic lymph-derived sequences at these positions. However, large differences in the length and number of
PNG sites between lymph node and brain V1/V2 loops were observed. Lymph node V1/V2 loops possessed an average of 70 amino acids and 6 PNG sites, while brain V1/V2 loops were more compact, with an average of 60 amino acids and 5 PNG sites. Together, these data suggest the major determinants of macrophage tropism in patient E21 are located in the V1/V2 loops.

Blood-derived gp160s from patient KM34 were generally non-macrophage-tropic or conferred low levels of macrophage tropism, while semen-derived gp160s were either highly macrophage-tropic or non-macrophage-tropic (Fig. 4). Despite differences in macrophage tropism, residue N283, CD4 contact residues, and other residues implicated in modulating macrophage tropism did not exhibit compartment specific differences, although some variability was observed. Inspection of the encoded gp160 proteins revealed the observed range of tropism in patient KM34 is likely to be modulated by differences in the V1/V2 region, where substantial differences in the length and number of PNG sites between blood- and semen-derived V1/V2 loops were observed (Fig. 5B). Blood V1/V2 loops possess an average of 66 amino acids with 6 PNG sites, while semen V1/V2 loops are more compact, with an average of 57 amino acids and 5 PNG sites (see Fig. 5B). Recombinant semen envelope KM34SeR63 had a blood-like V1 loop and exhibited a 50-fold reduction in macrophage tropism (0.76%) compared to semen-derived KM34SeR3 (recombinant, blood-like C3 region, 37.40%) and KM34SeR33 (nonrecombinant, 39.11%). Indeed, nonrecombinant seminal isolate KM34SeR33 is identical to recombinant KM34SeR63 at all gp120 residues outside the V1 loop (see Fig. 5B). These data
indicate a blood-like V1 loop in a seminal background renders gp160 KM34SeR63 non-macrophage-tropic and demonstrate cellular tropism modulation due to intercompartment env recombination.

Maraviroc sensitivity. The CCR5 antagonist maraviroc demonstrates potent antiviral activity against R5-tropic primary isolates derived from diverse clades (14). As expected, the X4 gp160s KM34BIR71 and E21lnD43 were highly resistant to this entry inhibitor (Table 3 and Fig. 6). In contrast all of the R5 gp160s tested exhibited sensitivity to maraviroc (50% inhibitory concentration [IC50] range, 0.08 to 1.33 nM), regardless of tissue origin. In patient E21, compartmentalization of sensitivity to maraviroc was observed. Nonrecombinant brain-derived gp160s were highly sensitive (mean IC50 0.18 nM), while lymph node R5 isolates exhibited 5-fold less sensitivity (mean IC50 1.06 nM). Sensitivity to maraviroc can be modulated by sequence differences in the V3 loop, which is the principle determinant for CCR5 coreceptor binding. Amino acid residues that differed between brain- and lymph-derived sequences were located at positions L309I, G315R, T319A, and so on.

### TABLE 2. Putative breakpoints of intercompartment recombinants

<table>
<thead>
<tr>
<th>Mosaic sequence</th>
<th>No. of breakpoints</th>
<th>Parental sequences</th>
<th>Region</th>
<th>No. of supporting sites</th>
<th>Breakpoint position</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>E21BrD46</td>
<td>1</td>
<td>E21BrD82</td>
<td>1–1956</td>
<td>22</td>
<td>2101</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>E21BrD92</td>
<td>1</td>
<td>E21LnD10</td>
<td>2161–2571</td>
<td>6</td>
<td>1846</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>E21BrD100</td>
<td>1</td>
<td>E21BrD62</td>
<td>1–1650</td>
<td>22</td>
<td>2210</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>E21BrD107</td>
<td>3</td>
<td>E21BrD90</td>
<td>1–253</td>
<td>2</td>
<td>262</td>
<td>0.0442</td>
</tr>
<tr>
<td>E21BrD107</td>
<td>3</td>
<td>E21LnD56</td>
<td>364–687</td>
<td>8</td>
<td>775</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>E21BrD107</td>
<td>3</td>
<td>E21BrD90</td>
<td>829–1392</td>
<td>16</td>
<td>1515</td>
<td>0.0056</td>
</tr>
<tr>
<td>E21BrD107</td>
<td>3</td>
<td>E21LnD56</td>
<td>1650–2571</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KM11BIR7</td>
<td>2</td>
<td>KM11BIR22</td>
<td>1–93</td>
<td>2</td>
<td>219</td>
<td>0.0365</td>
</tr>
<tr>
<td>KM11BIR7</td>
<td>2</td>
<td>KM11SeR2</td>
<td>394–848</td>
<td>9</td>
<td>864</td>
<td>0.0004</td>
</tr>
<tr>
<td>KM11BIR7</td>
<td>2</td>
<td>KM11BIR22</td>
<td>914–2571</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KM34SeR3</td>
<td>2</td>
<td>KM34SeR21</td>
<td>1–837</td>
<td>17</td>
<td>911</td>
<td>0.0004</td>
</tr>
<tr>
<td>KM34SeR3</td>
<td>2</td>
<td>KM34BIR71</td>
<td>1017–1080</td>
<td>4</td>
<td>1164</td>
<td>0.0569</td>
</tr>
<tr>
<td>KM34SeR3</td>
<td>2</td>
<td>KM34SeR21</td>
<td>1227–2571</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KM34SeR54</td>
<td>2</td>
<td>KM34SeR49</td>
<td>1–837</td>
<td>14</td>
<td>905</td>
<td>0.0007</td>
</tr>
<tr>
<td>KM34SeR54</td>
<td>2</td>
<td>KM34BIR71</td>
<td>1017–1080</td>
<td>4</td>
<td>1173</td>
<td>0.0286</td>
</tr>
<tr>
<td>KM34SeR63</td>
<td>2</td>
<td>KM34SeR26</td>
<td>1–238</td>
<td>3</td>
<td>319</td>
<td>0.0356</td>
</tr>
<tr>
<td>KM34SeR63</td>
<td>2</td>
<td>KM34BIR11</td>
<td>363–450</td>
<td>5</td>
<td>551</td>
<td>0.0010</td>
</tr>
<tr>
<td>KM34SeR63</td>
<td>2</td>
<td>KM34SeR26</td>
<td>557–2571</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The coordinates are given relative to the nucleotide position in the HXB2 reference strain env gene.

---

**FIG. 3.** Phylogenetic trees corresponding to regions of differential homology in intercompartment mosaic E21BrD107. Patient E21’s env alignment was divided into four segments, in agreement with the three breakpoints identified in isolate E21BrD107, and ML phylogenetic trees were constructed for each segment: positions 1 to 262 (A), positions 263 to 774 (B), positions 775 to 1514 (C), and positions 1515 to 2571 (D). Highlighted isolate E21BrD107 exhibits significant positional shifting between compartment specific clades.
and Q322E (Fig. 5A). Intercompartment mosaic E21BrD107 contains a lymph node-like V3 loop (see Fig. 2) and possesses lymph node-like sensitivity to maraviroc (IC_{50} = 1.19 nM, Table 3 and Fig. 6). Together, these data indicate compartment specific differences in sensitivity to maraviroc in patient E21 and demonstrate maraviroc sensitivity switching due to intercompartment env mosaicism. In patient KM34, there was limited variability in the V3 loop sequences of blood- or semen-derived isolates (Fig. 5B), and all of the R5 gp160s derived from this patient had similar maraviroc sensitivity.

**Soluble CD4 sensitivity.** Previously, we demonstrated that R5 macrophage tropism is significantly correlated with increased sensitivity to sCD4, which is compatible with increased exposure of the CD4 binding site and/or substitutions which confer tighter CD4 binding (49). The macrophage-tropic gp160s derived from brain and semen samples of patients E21 and KM34 were highly sensitive to entry inhibition by sCD4, whereas the non-macrophage-tropic gp160s derived from corresponding lymph node and blood were demonstrably less sensitive to sCD4 entry inhibition (Table 3). Intercompartment recombination had minimal effect on the sensitivity to sCD4 inhibition.

**Sensitivity to broadly neutralizing antibodies IgG1-b12 (b12) and 2G12.** The monoclonal antibody b12 binds to a conformational conserved epitope on gp120 overlapping a discrete subset of CD4 contact residues (91). Previous data has shown that subtype B macrophage-tropic brain-derived gp160s are more sensitive to b12 than non-macrophage-tropic lymph tissue-derived gp160s (17, 49). In complete contrast to these findings, the non-macrophage-tropic lymph-derived gp160s from patient E21 were highly sensitive to b12 neutralization, while brain gp160s exhibited b12 resistance (Table 3). Regulation of b12 sensitivity, albeit in CRF01_AE viruses, has been linked to two N-linked glycan sites at amino acid positions 186 and 197 in the V2 and C2 regions (78) (Fig. 5A). Mosaic envelope E21BrD107 is similar to lymph-derived gp120s across nearly all regions except the V1V2 region (Fig. 2) and yet maintains a brain-like b12-resistant phenotype. Together, these data suggest the determinants contributing to b12 resistance/sensitivity in patient E21 gp160s are located in one or more of the V1/V2/C2 regions. The gp160s from patient KM34 (non-subtype B) were largely resistant to b12 inhibition, irrespective of their macrophage tropism or source.

The monoclonal antibody 2G12 targets a carbohydrate

### Table 3. Phenotypic properties of patient-derived gp160s

| gp160 | Sample material | Coreceptor usage | Property
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Macrophage tropism</td>
</tr>
<tr>
<td>E21LnD4</td>
<td>Lymph node</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>E21LnD10</td>
<td>Lymph node</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>E21LnD17</td>
<td>Lymph node</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>E21LnD24</td>
<td>Lymph node</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>E21LnD43</td>
<td>Lymph node</td>
<td>CXCR4</td>
<td>–</td>
</tr>
<tr>
<td>E21LnD58</td>
<td>Lymph node</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>E21BrD2</td>
<td>Brain</td>
<td>CCR5</td>
<td>+</td>
</tr>
<tr>
<td>E21BrD9</td>
<td>Brain</td>
<td>CCR5</td>
<td>+</td>
</tr>
<tr>
<td>E21BrD37</td>
<td>Brain</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>E21BrD82</td>
<td>Brain</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>E21BrD62</td>
<td>Brain</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>E21BrD100 (R)</td>
<td>Brain</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>KM34BrR14</td>
<td>Blood</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>KM34BrR14</td>
<td>Blood</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>KM34BrR52</td>
<td>Blood</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>KM34BrR66</td>
<td>Blood</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>KM34BrR70</td>
<td>Blood</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>KM34BrR71</td>
<td>Blood</td>
<td>CXCR4</td>
<td>–</td>
</tr>
<tr>
<td>KM34CeR3 (R)</td>
<td>Semen</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>KM34CeR33</td>
<td>Semen</td>
<td>CCR5</td>
<td>++</td>
</tr>
<tr>
<td>KM34CeR53</td>
<td>Semen</td>
<td>CCR5</td>
<td>–</td>
</tr>
<tr>
<td>KM34CeR63 (R)</td>
<td>Semen</td>
<td>CCR5</td>
<td>–</td>
</tr>
</tbody>
</table>

* (R), characterized recombinant gp160.

b Macrophage tropism is scored as follows: –, <1%; +, 1 to 5%; ++, 5 to 15%; and ++++, >15% infectivity in HeLa TZM-BL cells. Maraviroc sensitivity (nM) is scored as follows: ++++, IC_{50} = <0.25; ++, IC_{50} = 0.25 to 1.0; +, IC_{50} = >1; and –, CXCR4 user. sCD4, b12, and 2G12 sensitivity (μg/ml) is scored as follows: ++++, IC_{50} = <10; ++, IC_{50} = 10 to 20; +, IC_{50} = 20 to 50; and –, >50.

**FIG. 4.** Macrophage tropism of HIV-1 gp160s amplified from anatomically discrete tissues. Patient/tissue-specific gp160s were expressed on HIV-1 pseudovirions. env’ pseudovirions were titrated on HeLa TZM-BL cells and on primary macrophages. Macrophage infectivity is expressed as a percentage of HeLa TZM-BL cell infectivity.
epitope, primarily in subtype B viruses, composed of high-mannose/hybrid glycans in the C2, C3, C4, and V4 domains (64). Lymph node-derived gp160s from patient E21 were susceptible to 2G12 neutralization, while brain-derived gp160s were resistant (Table 3). PNG sites at N295, N332, N339, N386, N392, and N448 are reportedly important for 2G12 binding (38, 49, 64, 66). The PNG site at position 332 is absent from E21 brain-derived gp160s but present in lymph node-derived gp160s. However, the brain-derived mosaic E21BrD107 has a PNG site at position N332 and yet a 2G12-resistant phenotype, suggesting that 2G12 resistance exhibited by E21 brain-derived gp160s is associated with determinants other than those previously described. Indeed, lymph node gp160s contain an extra PNG site in the V2 loop between N186 and N197 (see Fig. 5A), which may confer susceptibility to 2G12.
in vitro artifacts which directly affect datasets generated and the subsequent interpretation of patterns of sequence evolution have utilized bulk amplification, cloning and sequencing strategies, or have utilized viruses selected for growth in vitro. These methods are subject to numerous biases and in vitro artifacts which directly affect datasets generated and the subsequent interpretation of patterns of sequence evolution (62, 73, 80, 81). In addition, compartmentalization studies often utilize relatively small fragments of env (V1-V3 or C2-V5), reducing the chances of detecting putative recombination events in highly similar sequences. To minimize any experimentally induced biases and maximize the phylogenetic signal apparent in our data set, as well as facilitating phenotypic analyses, we utilized full-length env sequences amplified from single molecule templates derived directly from paired patient tissue samples. This methodology has been demonstrated to minimize polymerase induced misincorporations, reducing artificial inflation of intrahost viral diversity. If introduced in the initial rounds of amplification, the signatures of polymerase induced errors are easily identifiable since they appear as dual peaks in the sequence chromatographs. If introduced in the later rounds of amplification, misincorporations remain undetected since they represent only a minor component of generated amplicons (62). Critically, viral sequences generated in this manner are not subject to in vitro recombination events, which occur via polymerase template switching during bulk amplification from multiple starting templates (42, 62, 70).

Numerous phylogenetic and statistical tests can be used to assess degrees of HIV-1 intrahost compartmentalization. However, genetic mosaicism has been shown to adversely affect the results of these tests (90). Furthermore, identifying recombination events in datasets derived from highly related variants is challenging. Automated procedures for the detection of recombination implemented in GARD (28, 29) identified numerous significant breakpoints in all datasets, as defined by discordant topologies, but lacked the facility to identify parental strains requiring subsequent manual investigation of the sequence data. Intercompartment mosaics breakpoints and parental strains could be correctly detected in via tests used in RDP2 (40). However, disparate results were obtained using the various tests used in this program, which required some prior knowledge of the likely recombinants and parental strains for correct interpretation of the output data. Consequently, we utilized an iterative approach involving manual inspection for topological incongruence in trees derived from different regions of env. Pairwise diversity plots and informative site arrays then enabled high-resolution identification of breakpoints and parental sequences. A number of breakpoints within env were identified across the intercompartment recombinant forms we characterized. Although specific breakpoint hot spots were not obvious, breakpoint locations generally occurred in the more conserved regions of env or were located at gp120 conserved/variable region boundaries. Breakpoint locations were generally in agreement with intersubtype breakpoint hot spots identified by in vitro modeling and in vivo observations (71). This is despite the fact that predicted constraints, such as maintenance of protein fold, are likely to be less severe at the intrahost level compared to the intersubtype level: viral variants circulating within a host infected with a single strain possess limited diversity while different group M subtypes can differ by up to 38% in envelope amino acid identity (82). All of the recombinant gp160s tested were shown to be functional in pseudotype assays, indicating that the characterized recombination events resulted in no deleterious disruption of gp160 tertiary or quaternary conformation and stability. In addition, it is likely that genomic constraints may also influence patterns of HIV-1 env recombination. Secondary
RNA structures apparent in the env gene region (84) may favor/inhibit certain recombination events in naturally circulat ing virus. Variable loop regions (V1 to V5) in gp120 are reported to contain largely unstructured genomic RNA but are bordered by evolutionary conserved RNA structures in the conserved regions (84). Our identified breakpoints were generally observed in conserved regions or at conserved/variable region boundaries. Mosaic genomes that result in disruption to gp160 protein conformation or evolutionary conserved genomic RNA elements will be deleterious and removed from populations via purifying selection.

Viruses circulating in the peripheral blood or present in lymph nodes evolve to occlude critical regions, such as receptor binding sites, from host neutralizing antibodies. The blood-brain barrier and blood-testes barrier restrict the passage of neutralizing antibodies, making the brain and testes immunoprivileged tissues (5). Similarly, the type and frequency of HIV-1 permissive cells differs between sites, and the relative availability of CD4 and coreceptors on these target cells is likely to influence their permissiveness for different strains (49, 50). Consequently, viruses replicating in the brain or semen are likely to evolve in response to their unique environment, acquiring tropism and receptor affinity appropriate to each niche. Our analyses show that compartmentalization can be associated with altered macrophage tropism and sensitivity to entry inhibitors and monoclonal neutralizing antibodies.Crudially, our data has shown that these phenotypes can be altered by intercompartment recombination events. For example, recombinant brain isolate E21BrD107, which possesses a lymph node-like C2-C5 region, maintained the macrophage tropism of nonrecombinant brain-derived gp160s but acquired a reduced sensitivity to maraviroc through acquisition of V3 loop represented in lymph node-derived gp160. In addition, in patient KM34, blood-derived viruses were non-macrophage-tropic or poorly macrophage-tropic, whereas semen-derived gp160s were highly macrophage-tropic. However, recombinant semen envelope KM34SeR63, which possessed a region of gp120 encompassing the V1 loop present in blood-derived gp160s, was non-macrophage-tropic. At the population level, the wide global dissemination of multiple intra- and intersubtype CRFs indicates recombinant genomes may possess a fitness advantage compared to pure subtypes in certain environmental settings. At the intrahost level, intercompartment recombination represents an additional molecular mechanism for generating viable and potentially fitter viral variants. Our study reveals the genetic signatures of intercompartment mosaicism are readily detectable in viruses isolated from different tissues, indicating this process occurs frequently in infected individuals. Indeed, although the present study was on a relatively small scale, with five patient’s viral populations investigated, intercompartment mosaicism was detected in three individuals, with recombinant env genes comprising between 2 to 11% of generated amplicons. Our study was limited to viruses amplified from two tissue compartments per patient derived from a single time point, and characterization of recombinant isolates was restricted to those where both parental strains were represented in the population. Consequently, as HIV-1 is reported to be present in multiple tissue types, it is likely that intercompartment recombination occurs at rates in excess of those reported herein. Investigation of the functional consequences of the shuffling of phenotypic traits suggests this process may facilitate viral immune evasion and development of resistance to entry inhibitors. Viral progeny generated in this manner possessing advantageous combinations of traits will be rapidly swept to fixation within an infected host.

ACKNOWLEDGMENTS

We thank Gerry Gillean for indispensable help with sample acquisition. This study was supported by NIH grant R01HD049273-02.

REFERENCES