AIDS Rev 2002;4:195-212

HIV-1 Genetic Recombination: Experimental Approaches and Observations

Wenfeng An and Alice Telesnitsky

Great Lakes Center for AIDS Research, Department of Microbiology and Immunology, University of Michigan Medical School, Michigan, USA

Abstract

Many HIV-1 isolates contain interwoven genomic sections derived from multiple parental strains. Such chimeric genomes arise via genetic recombination. This review summarizes experimental approaches for addressing the frequency of HIV-1 genetic recombination during single cycles of viral replication *in vitro*, and describes factors –such as variation in extents of sequence homology and the metabolic state of the infected cell– that modulate recombination. Findings from such studies suggest that recombinogenic template switching is an even more common occurrence during HIV-1 DNA synthesis than is the introduction of base substitution errors. This implies that recombination is an inherent property of retroviral DNA synthesis, and that the vast majority of HIV-1 DNAs are biochemical recombinants.

Key words

HIV. Recombination.

Introduction

Early molecular characterization revealed that HIV-1 strains could be assigned to genetically distinct clades. These clades, now known as subtypes, differ from one another in genetic variation that likely arose via sequential mutations in genetically isolated branches of HIV-1 after its introduction into humans[66]. Once sequencing full-length HIV-1 genomes became common practice, it became apparent that the ancestry of many viral isolates was not linear. Instead, these recombinant strains contained patches of sequence derived from more than one previously recognized subtype[65,90].

Genetic recombination in animal retroviruses had been known for decades, and thus the potential for recombination to contribute to HIV-1 diversity was recognized as soon as it was discovered that AIDS is caused by a retrovirus. In fact, early work with avian RNA tumor viruses suggested that recombination was so frequent that genetic markers re-assorted essentially as if they were unlinked[120]. This is surprising when one considers that the entire genetic content of retroviruses is contained on a single RNA[114]. Because covalently linked genes cannot re-assort on separate RNAs as they do for multi-segmented viruses such as influenza, retroviruses must have evolved a mechanism to re-assort physically linked genes at an unprecedentedly high frequency[19,62].

Early work to determine whether or not HIV-1 and other primate lentiviruses could recombine demonstrated that genetic recombination was readily detectable. For example, Clavel, et al. showed that an integrated provirus containing a

Correspondence to:
Alice Telesnitsky
Great Lakes Center for AIDS Research
Department of Microbiology and Immunology
University of Michigan Medical School
1150 West Medical Center Drive Room 5641
Ann Arbor, MI 48109-0620 USA
E-mail: ateles@umich.edu

stop codon could be rescued by recombination with a second defective virus that had a different mutation[16]. Moutouh, Corbeil and Richman, as well as Kellam and Larder, found that recombination could lead to the co-segregation of drug-resistance markers in tissue culture[51,70], and Wooley, et al. demonstrated that SIV recombination occurred in an experimentally co-infected rhesus monkey[118]. However, because it is difficult to determine generation times in a replicating virus population, and because competitive advantage largely dictates population composition[20], it would be inaccurate to deduce the frequency of recombination from the prevalence of replication-competent recombinants.

This review focuses on experimental work that has examined, more directly, the frequency of HIV-1 genetic recombination during single cycles of replication, and on factors that modulate recombination under experimentally controlled conditions. Until quite recently, most such work was performed in simple retrovirus systems, and thus much of the technology and findings described here are rooted in studies with alpha- and gamma-retroviruses. Studies with HIV-1 reveal both similarities and differences between genetic recombination for HIV-1 and for simple retroviruses.

Mechanism: reverse transcriptase template switching between co-packaged RNAs

Most retroviral recombination results from reverse transcriptase (RT) copying part of one viral genome and then switching to a homologous region of a second genome to complete DNA synthesis[18]. In the example shown in figure 1, one parental RNA contains a protease (PR) allele with mutations that confer broad resistance to protease inhibitors (resistance is indicated by the superscript ^{R}), while the second parental RNA contains RT-coding sequences that confer resistance to AZT. Because both high level resistance to AZT and cross-resistance to protease inhibitors require multiple alterations to wild-type sequences, developing either form of resistance can take several viral generations[28]. In contrast, once each resistance has developed independently, recombination between genomes permits co-segregation of both traits in a single cycle of replication.

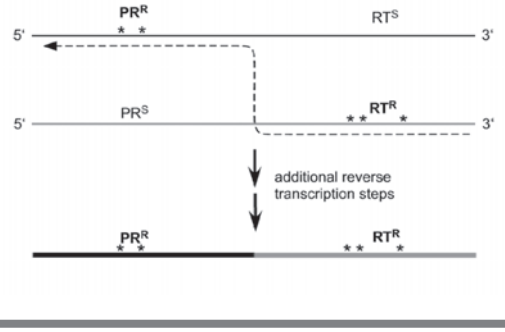

Figure 1. *Recombinogenic template switching during reverse transcription. Thin solid lines represent genomic RNAs, dotted line indicates nascent DNA. The direction of DNA synthesis that switches between RNA templates is shown by an arrowhead, and the thick solid line represents the final double-stranded DNA product. * indicates sites of mutations responsible for drug-resistance phenotypes. Note that although all protease-resistance mutations indicated in this schematic diagram reside in the PR coding region, clinically-relevant secondary mutations have also been observed at other locations such as at protease cleavage sites[28].*

A critical factor that contributes to HIV-1 recombination frequency is the unusual genome organization of retroviruses. Although each virion RNA includes the entire genome, retroviruses differ from other viruses in that each contains duplicate RNAs[114]. Retroviruses are sometimes considered to be diploid, and virions that contain two different RNAs are described as heterozygous in this review. Co-packaging two complete RNAs in a single virion provides two templates to the reverse transcription machinery and is a critical factor in the high frequency of retroviral recombination.

Another important factor is the unusual properties of RT. Most DNA polymerases do not readily perform the sorts of template switches that are required for retroviral recombination. It has been postulated that the reason reverse transcription is prone to recombinogenic switching is because replicative template switches are required during the synthesis of every viral DNA (Fig. 2)[21,111]. During every round of reverse transcription, viral genomic RNA templates a double-stranded DNA product that is longer on both of its ends than is the template RNA. As outlined in figure 2, the so-called "jumps", or strong-stop template switches during reverse transcription result in the joining and duplication of sequences found only once in genomic RNA, thereby generating the long terminal repeats (LTRs)[29]. Replicative strong-stop template switching and recombination are believed to be mechanistically similar processes[111]. Whenever template switches in addition to strong-stop jumps occur, the product DNA can be considered a biochemical recombinant. Genetic recombination —re-assortment of genetic markers— can result when recombinogenic template switching occurs between templates that differ from one another.

Ongoing research is addressing which properties of the reverse transcription machinery contribute to its recombinogenic properties. One implicated factor is the viral nucleocapsid protein (NC), which binds retroviral RNA and functions as a nucleic acid chaperone[93]. NC affects template switching in *in vitro* reconstituted reaction studies, and its effects on recombination during replication are currently under examination[73,131]. RT itself is also a likely candidate. All retroviral RTs contain two active sites: one for DNA polymerization and a second for RNaseH, the nuclease that degrades RNA from RNA-DNA duplexes[109]. Structure-function studies have revealed that some alterations to

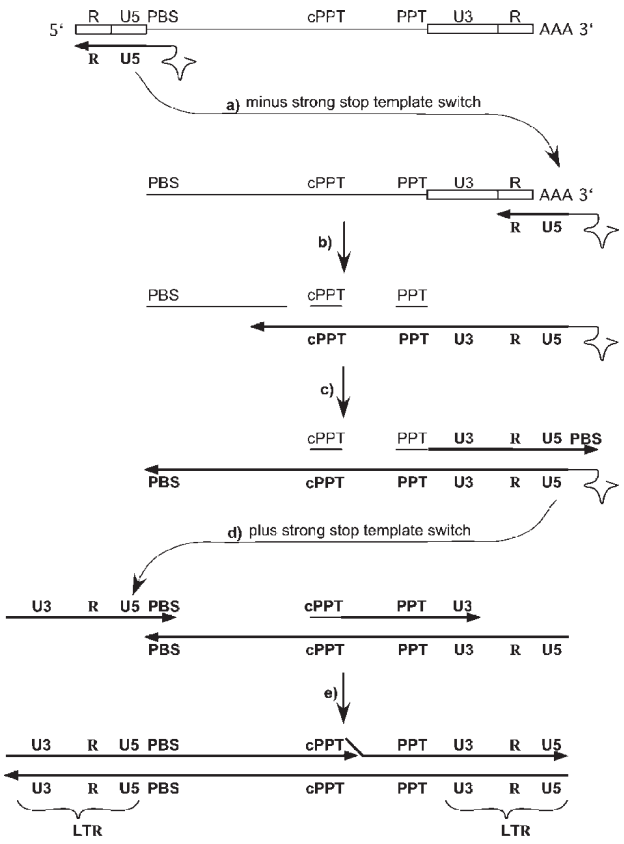

Figure 2. *Replicative template switches during HIV reverse transcription. Outline of prevailing model for retroviral DNA synthesis[29] modified to reflect the presence of a secondary plus-strand DNA initiation site (the central polypurine tract [cPPT][14,15]) in HIV and other lentiviruses. a) The initial minus-strand DNA intermediate synthesized at the RNA's 5' end is transferred to the 3' end during minus strong-stop template switch. b) Continued minus-strand synthesis is accompanied by RNaseH degradation of the RNA template. c) Residual polypurine tract (PPT) oligoribonucleotides prime plus-strand DNA synthesis. d) The plus-strand product that results from synthesis into the tRNA primer is transferred to the 5' end of nascent minus-strand DNA during plus-strand strong-stop transfer. e) Completion of both DNA strands results in a gapped duplex with long terminal repeats (LTRs) on either end.*

gammaretrovirus RT affect template switching[85,104], but these studies have not yielded a clear picture of RT structural determinants required for recombination.

One property of RT —its RNaseH activity— is an undisputed requirement for high frequency template switching[11,13]. Template switching frequencies are decreased several fold for gammaretroviruses that contain certain RNaseH-defective forms of RT or that have reduced levels of active RNaseH[10,85,104]. These findings are consistent with "minus-strand exchange"[10,18,73,84] (Fig. 3), a currently favored model for recombination during minus-strand synthesis.

RNaseH activity is required for HIV replication[98,112], but proviral DNA can be formed by virions with less than the wild-type complement of RNaseH[45]. This observation, paired with findings that demonstrate limiting RNaseH increases recombination[10], suggests the hypothesis that retroviruses with relatively low RNaseH activity, such as HIV-2[38], may recombine less frequently than HIV-1, which possesses relatively high RNaseH activity. It should be noted, however, that naturally arising recombinants have been reported, even for HIV-2[64,107].

Note that the model presented in figure 3 describes recombination during minus-strand synthesis. Models for plus-strand recombination have also been proposed[48,49]. Available experimental evidence suggests that most retroviral recombination occurs during minus-strand synthesis[3,127]. However, an important caveat to these studies is that all have been performed with gammaretroviruses, which differ from lentiviruses and alpharetroviruses in that they use a single initiation site for plus-strand synthesis[15,29,55]. Additionally, as summarized below, RNA packaging properties appear to differ between gammaretroviruses and lentiviruses, and thus gammaretrovirus genetic observations can misrepresent biochemical processes. Plus-strand recombination may indeed prove to be more frequent for viruses such as HIV-1, which synthesize plus-strands discontinuously, than they appear to be for gammaretroviruses. Con-

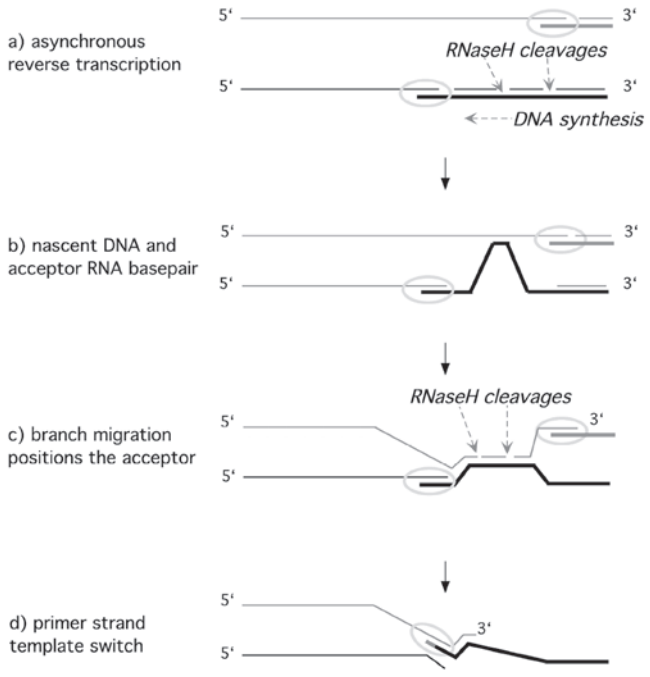

Figure 3. *Minus-strand exchange model for genetic recombination. a) DNA synthesis becomes asynchronous once RT advances further on one co-packaged RNA (the donor template) than on the other. b) Cleavage of the RNA template strand by RNaseH exposes single stranded minus sense DNA, which can base-pair with complementary sequences on the acceptor RNA. c) Branch migration of the nascent DNA: acceptor RNA duplex positions the donor template growing point near the acceptor template, and RNaseH degradation of the donor template causes the slower RT complex to abort. d) Template switch is accomplished by a repositioning of the primer strand onto the acceptor template, as has been demonstrated in reconstituted reactions[95]. Thin lines represent RNA, thick lines represent DNA, ovals represent RT.*

sistent with this possibility are the observed "backwards" sequence insertions, suggestive of plus-strand recombination, which have been reported for certain retro-elements and alpharetroviruses[78,105]. Recombination within proviral DNA has also been postulated to contribute to retroviral genetic variation, and phylogenetic comparisons of integrated retro-elements support the existence of variation generated at the DNA level. However, it seems likely that some early experimental approaches, used to suggest DNA-level recombination, were subject to transfection-related artifacts[30,31]. More recent experimental work suggests that DNA-mediated retroviral recombination is far less frequent than recombination which occurs during reverse transcription, and is likely no more frequent than other forms of cellular DNA recombination ([58] and references therein).

Single cycle of replication assays

Single replication cycle assays are the preferred way to study recombination, because events can be detected at their actual biochemical frequency rather than allowing selection for products with a competitive advantage, as would occur during a spreading infection[123]. Single cycle assays use vector and helper systems, like those used to generate lentiviral vectors for gene transfer[68]. One assay component is a reporter vector that contains one or more selectable marker genes embedded in sequences that include *cis*-acting signals such as the packaging signal, Ψ, and those required for reverse transcription. The second component is the *trans*-acting factors required to produce virions. These can either be provided by packaging cells that express virion proteins, or by co-expressing vector and packaging defective (Ψ-) virion protein expression constructs in producer cells. Virions harvested from producer cells are applied to uninfected target cells. Vector RNAs packaged in these virions become reverse transcribed and integrated into target cell DNA, and the resulting proviruses are maintained as stable genetic elements. Even rare reverse transcription events can be detected if vectors are engineered to contain selectable markers, because cells with vector proviruses can be selected and amplified by cell division in the same way that *E. coli* – containing plasmids can be selected on antibiotic-containing agar. A spreading infection does not result if the vectors do not encode essential viral proteins, and thus replication is limited to a single cycle. Typically, vector provi-

rus-containing cells are clonally expanded, and structures of integrated proviruses are analyzed by standard molecular approaches.

Gammaretrovirus two vector recombination assays

Pioneering studies by Howard Temin, et al. examined recombination between co-packaged gammaretrovirus vectors[40,41,111]. An overview of these approaches is presented in figure 4. Based on these researchers' confirmation of earlier observations[121] that recombination is not observed when two viruses co-infect a single cell but is detected among products of virions from dually infected cells[40], an initial step in these assays was to establish heterozygous virion producer cell lines. These cells co-expressed virion proteins and two different retroviral vectors. The vectors were highly similar, but each was engineered to express only one or the other of two different drug-resistance markers[40]. These vectors also differed at other positions (indicated by * in figure 4a), to allow genetic screening for recombination in vector intervals other than the target region between selectable markers.

Because virion producer cells co-expressed both vectors, some of the resulting virions contained both RNAs. If recombination between marker-inactivating mutations occurred during reverse transcription of such heterozygous virions, a provirus encoding resistance to both drugs could result. Provided that the multiplicity of infection was low enough that the chance of single cell co-infection was minimal, the number of target cells that expressed both markers was an indication of genetic recombination.

Recombination frequencies can be calculated from single- and double-resistant selectable marker titer data after consideration of a number of parameters[123]. One is to calculate the predicted "recombining population" size, or fraction of virions that are capable of generating detectable recombinants because they have co-packaged both vectors. It has been assumed that if two packaga-

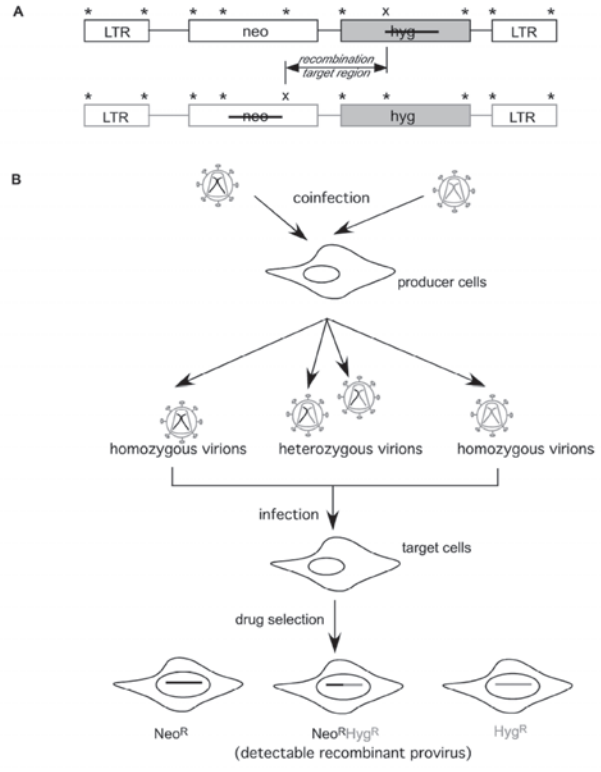

Figure 4. *Two-vector recombination assays. a) Schematic representation of gammaretrovirus recombination assay vectors used by Hu and Temin[40,41]. * indicates position of genetic difference between templates; x indicates genetic difference responsible for inactivating marker gene; recombination target region is interval between inactivating mutations. LTR indicates long terminal repeat, which contains U3, R, and U5 sequences (see figure 2), neo and hyg indicate drug-selectable marker genes. Crossed-out text represents inactivated gene. b) Overview of experimental scheme for recombination assays that use two vectors such as those in figure 4a. Two different proviruses are introduced into single producer cells: in the example here, by co-infection. Virions are then harvested from vector co-expressing producer cells. As described in the text, some virions are predicted to contain RNA heterodimers. Vector-containing virions are used to infect target cells. Progeny provirus-containing cells are identified by drug selection. Provided each parental vector conferred resistance to only one drug, and infections were performed at a low multiplicity of infection, dual drug resistance is diagnostic of recombination. Note that in this figure, virions are represented with capsid morphology more reminiscent of HIV-1 than of gammaretroviruses.*

ble RNAs are co-expressed, they will associate and dimerize randomly, based on their intracellular proportions[19]. This generates heterozygous and homozygous virions at frequencies predicted by the equation

$$A^2 + 2AB + B^2 = 1$$

which is known in population genetics as the Hardy Weinberg equation. When the concentration of one RNA, 'RNA A', is the same as that of RNA B, then AB heterodimers should be present in half of the total virions, and AA and BB homodimers are predicted to be present in 25% of the total apiece. For recombination frequency calculations, the proportions of co-expressed RNAs in one dimer form or another have often been inferred from single drug resistance titers[41,124].

To calculate recombination frequencies, dual drug resistance titers per calculated recombining population member are doubled to account for the fact that for every dual resistant provirus, the reciprocal (an undetected dual-drug sensitive) recombinant should form[44,129]. Values are also adjusted to account for the fraction of the genome length in which recombination was measured (the recombination target), which in the figure 4 example is the distance between drug resistance-inactivating mutations. If the target interval were 0.8 kb and the entire vector were 8 kb, then target interval recombination values must be multiplied by 10 to yield a genome-wide rate. These approaches have yielded frequencies of approximately one recombination event during the synthesis of every two to seven gammaretroviruses[4,40,41,43].

Potential weaknesses of two-vector recombination assays include the fact that they are technically challenging and that they are based, in part, on some theoretical assumptions. One assumption is that recombination occurs at all genomic positions at more-or-less uniform frequencies rather than principally at specific sites. Experimental work has confirmed that although recombination frequencies are not uniform –indicating the presence of "hot" and/or "cold" spots for recombination– recombination is observed throughout the HIV-1 genome[3,42,117]. Recombinogenic crossover frequencies are at least roughly proportional to recombination target length for both gammaretroviruses and HIV-1[2,46,130].

One technical challenge is to establish a situation where the investigator can be reasonably certain whether or not a particular proviral product was generated by a heterozygous virion. One way that this has been addressed is to co-express vector RNAs at very different levels[79]. If 99% are RNA A and only 1% are RNA B, then Hardy Weinberg predicts that the vast majority of proviral products harboring marker B were derived from AB heterozygotes. These issues are addressed in the classic series of HIV-1 genetic recombination experiments outlined below.

Dougherty and Preston HIV-1 recombination experiments

Dougherty, et al. were the first to report single replication cycle two-vector recombination assays for HIV-1[42,124]. Assay vectors were derived from two HIV-1 strains, HXB2 and BCSG3, respectively, which differ in nucleotide sequence by approximately 5%. Both vectors retained most of the native HIV-1 genomic sequences except that part of *env* was replaced by a different marker gene in each vector: that for xanthine-guanine phosphoribosyltransferase (gpt) in the HXB2-based vector, and the puromycin (puro) resistance gene in the vector derived from BCSG3 (Fig. 5). Thus, the vectors themselves provided all replication proteins except *env*. These vectors were individually pseudotyped and introduced by sequential infection into a CD4-negative producer cell line that contained inducible HIV-1 *env*, so that infectious vector-containing virions were produced upon induction. Virions were harvested and used to infect CD4-positive target cells. In this system, vector virus could not reinfect the producer cells, because those cells lacked CD4, and target cells

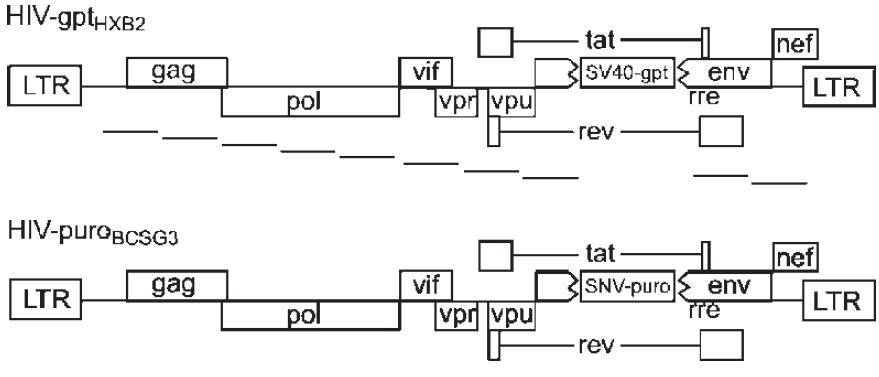

Figure 5. *HIV-1 based two-vector recombination assay vectors. See text.*

did not produce infectious virions, because vector provirus-containing target cells did not express *env*. Thus, replication was confined to a single cycle.

Virions from the producer cell clones that were chosen for further study generated gpt titers in target cells that were 100-fold higher than their puro titers. As discussed earlier, the Hardy Weinberg equation predicts that 99% of the puromycin-resistant proviruses synthesized by such virions will be generated by heterozygotes. Thus, by analyzing puromycin-resistant proviruses generated by virions from cells that expressed skewed ratios of vector RNAs, the authors could assume that the majority were products of heterozygous virions. A total of 86 separate puromycin-resistant proviral cell clones were isolated. Proviral DNA was amplified by PCR and various non-overlapping segments of the vectors (indicated by the bars under HIVgptHXB2 in figure 5) were subjected to heteroduplex tracking assay (HTA), using probes derived from the BCSG3-based vector. Because the two strains contained scattered differences throughout their lengths, homo- and hetero-duplex fragments yielded distinct gel migration patterns that were used to identify the parentage of each segment, and the genome intervals in which crossovers occurred were determined. The data yielded an average HIV-1 recombination rate of about three crossovers per genome per replication cycle, which was roughly 10-fold higher than those which had previously been reported for simple retroviruses, using similar experimental approaches[4,41].

Fidelity of genetic recombination

Substitution errors arise approximately once every 10,000 to 100,000 bases during viral DNA synthesis, or roughly once during the synthesis of one to ten genomes[89]. In reconstituted *in vitro* reactions designed to mimic HIV-1 template switching, base substitutions are observed at a significant percentage of all template-switching junctions[26,83]. These observations with purified RT led to the hypothesis that genetic recombination might cause many of the substitutions found in HIV-1 genomes[80,84]. It was suggested that transfer-associated errors could result from non-templated addition of an uncoded base followed by mispair extension upon growing point transfer to the second template. Non-templated addition occurs at more than 50% of all template ends when studied with purified RT[80,84]. Potential support for this possibility came from observations that errors at strong-stop switch junctions appear as frequently as in 5% of all replication products for some retroviruses and retroelements[53,71]. It has been suggested that host DNA recombination may be mutagenic[56], and the hypermutable recombination junctions that are observed for immune system gene rearrangements seem to confirm that some forms of recombination can be mutagenic[52].

However, in contrast to products from purified biochemical reactions, recombination junctions generated during replication in tissue culture appear fairly free of errors. Two-vector recombination assays have been used to study whether or not misincorporation errors occur more frequently at recombination junctions than during synthesis that proceeds without recombination[86,130]. In one study that addressed gammaretrovirus recombination fidelity, the entire recombination target region was sequenced from 29 individual products of two-vector recombination assays[130]. The frequency with which substitutions were observed —no substitutions in a total of about 1.5 kb sequenced— suggests substitutions at recombination junctions are not significantly more frequent than reported ~0.1 to 1 per 10 kb whole genome substitution frequencies[130]. Similar studies performed with HIV-1-based vectors concluded that HIV-1 recombination also is not particularly error-prone[133]. An alternate approach for studying gammaretrovirus fidelity involved a genetic system for forced recombinogenic template switching between the 5' end of one RNA and an acceptor target region on a co-packaged RNA. If template switching were faithful, a restriction site would form at the transfer junction, but if a non-templated acceptor template non-complementary base were added prior to template switching, the restriction site would be destroyed. The findings demonstrated that errors at that position occurred below the <1% threshold of detection[86].

Interestingly, in experiments where gammaretrovirus template switching frequency was suppressed by RNaseH limitation, the fidelity of DNA synthesis was drastically reduced[10]. Some or all of the reduced fidelity may have resulted from RT defects unrelated to template switching[132]. Nonetheless, template switching generally may be more of a fidelity factor than a mechanism of error introduction. During the reverse transcription of every retroviral DNA, RT must switch templates when it reaches its tRNA template's first modified base, which is perceived to be non-coding[106]. By analogy, it is possible that recombinogenic template switching may provide an error-escape mechanism when the reverse transcription machinery reaches a dysfunctional, environmentally modified base[25]. Error-avoidance —particularly from broken templates— is part of "forced copy choice" models for genetic recombination[21]. These models remain useful, even though it is now clear that factors unlikely to affect RNA integrity, such as nucleotide availability and genetic properties of RT, can modulate recombination, and that recombination rates appear to be higher than broken RNA frequencies[2,85,104,122].

An alternate way of studying recombination: repeat deletion

Two-vector recombination assays are prone to inaccuracies when experimental shortcuts are tak-

en, such as simultaneous rather than sequential co-transfection of vectors to generate producer lines[76]. Although such experimental "noise" can sometimes be circumvented, and is often tolerably within the range of experimental variation, alternate, more experimentally amenable ways of studying template switching are often preferable. Repeat deletion is one such method for measuring recombinogenic template switching.

This approach is based on observations that repeated sequences are often precisely deleted from retroviral vectors, suggesting that their removal is a form of homologous recombination[94]. In assays designed to exploit this, repeat deletion has been used to inactivate or reconstitute marker genes[46,126]. Figure 6a shows HIVlaacPuro, a repeat deletion vector that contains both a selectable marker (a puromycin resistance gene) and a screenable marker (*lacZ*)[1]. This vector was engineered to contain a coding region-internal sequence duplication that disrupts the LacZ translational reading frame, so that parental-form proviruses confer puromycin resistance, but do not express functional β-galactosidase. A functional *lac*Z gene is generated if precise repeat deletion occurs during reverse transcription, and cells harboring a deleted provirus stain blue when incubated with the chromogenic substrate, X-gal[87]. Because repeat deletion is presumed to be a recombination-like property, the blue to total (blue plus unstained) puromycin resistance titer is an indication of recombinogenic template switching frequency. Similar vectors that express *gfp* and can be scored by florescence-activated cell sorting have also been developed[104,126].

A major advantage of repeat deletion assays is that template switching can be monitored using a single vector. Because each vector contains both the sequence from which RT will switch (the donor template region) and the sequence to which RT will switch (the acceptor region), this approach obviates the need to determine template RNA ratios. Although repeat deletion is sometimes referred to as intramolecular recombination, a weakness of this experimental approach is that it is

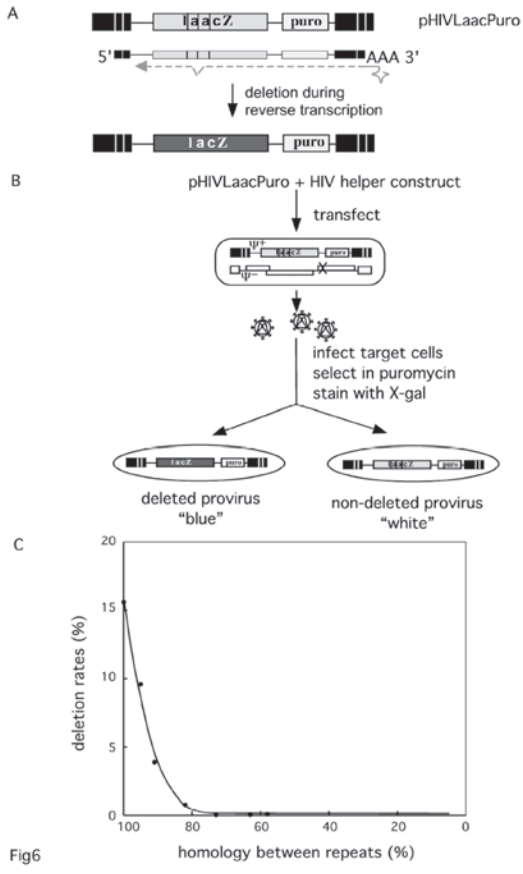

Figure 6. *repeat deletion assays and their application to assessment of homologous recombination. a) Schematic representation of HIVLaacPuro vector and repeat deletion during reverse transcription. Black boxes represent LTR elements; the internally repeated segments of lacZ are represented by repeated 'a'. b) Experimental overview, as explained in the text. Ψ+ and Ψ – refer to the fact that the vector contains a packaging signal but the helper function construct does not. X over env indicated its mutational inactivation in the helper that is represented. c) Effects of varying sequence similarity on the deletion frequency of 156 base repeats in vectors derived from HIVLaacPuro. Identical repeat deletion frequency was 14.9%.*

unclear whether or not repeat deletion is a good mechanistic mimic for genetic recombination. Possible concern is raised by findings from *E. coli*, where direct repeat deletion is RecA-independent but other forms of homologous recombination are RecA-dependent[61]. However, HIV-based assay data that compared deletion frequencies using a single vector RNA to switching rates between repeats located on two different vectors, suggest that repeat deletion does reproduce properties associated with retroviral intermolecular recombination[77]. This data was derived from assays that compared frequency of repeat deletion using HIVlaacPuro (Fig. 6) to the frequency of intermolecular switching between 'reciprocal repeat' vectors. Repeats in these later HIVlaacPuro derivatives differed by the introduction of a sufficient number of synonymous substitutions —27% differences overall— so that deletion between the two repeats on a single vector did not occur detectably. The vectors were designed so that one vector had the mutant repeat downstream of the wild-type repeat while the reciprocal vector contained the wild-type repeat downstream of the mutant sequence. Experimental values were adjusted to reflect the predicted heterozygous virion concentration and the assumption that while copying the 156 base repeat, the elongating polymerase would only be able to use one of the three potential acceptor template regions (the one on the same RNA and the two on the co-packaged RNA). The data indicated that switching between repeats occurred at a similar frequency, regardless of whether the acceptor repeat was on the same or on a co-packaged RNA[77].

Homologous vs non-homologous recombination

Work described thus far has focused on recombination between highly similar sequences. Retroviruses also perform non-homologous recombination between unrelated sequences. For example, the junctions between host and viral sequences in retroviruses with host-derived oncogenes presumably arose via non-homologous recombination (Fig. 7a)[128]. Gammaretrovirus assays designed to compare frequencies of homologous and non-homologous recombination concluded that non-homologous recombination was less frequent by two to three orders of magnitude[129].

Of interest to HIV biologists is how recombination is affected by sequence variation that lies

Figure 7. *Retroviral transduction. a) A model for host oncogene capture. Chance integration of a provirus upstream of a host oncogene allows occasional production of a read-through transcript that initiates in the upstream LTR and proceeds through the polyadenylation signal downstream of the oncogene. Once this RNA has been packaged and transmitted to a fresh cell, non-homologous template switching between sequences in the virus body (between LTRs) and sequences in appended downstream oncogene sequence can embed the oncogene in viral sequences. Note that other models for transduction have also been proposed, including ones in which rare host RNA co-packaging is envisioned to occur without physical linkage of host and viral RNAs. b) Transduction type recombination assay vectors. Δ117 under lacZ indicates that this allele contains a 117 base deletion; Δ134 and Δ290 in the 3' region indicate that downstream lacZ lacks 134 nucleotides of coding sequence on its 5' end and 290 on its 3' end. ΔpA indicates inactivated polyadenylation signal that results in read-through for all RNAs[77]. Note similarity in structure to postulated transduction intermediates in figure 7a. Whereas template switching is depicted as occurring from the 3' tail of one RNA to the body of the co-packaged RNA in both panels A and B, note that intramolecular switching between two portions of a single RNA may occur also or alternatively in either panel.*

between homology and non-homology. Recombination between related but differing sequences, which has occasionally been referred to as "homeologus" recombination, has been examined experimentally using a variation of the repeat deletion assay introduced in figure 6[1]. Synonymous substitutions were introduced into one copy of a repeated sequence in *lac*Z, while the native sequence was maintained in the second repeat. Identical repeat deletion frequencies were compared to rates for repeats that differed from one another by 5, 9, 18, 27, 37, or 42%. As presented in figure 6c, decreases in deletion rate with increasing genetic distances could be fit to an exponential decay curve. As little as 5% difference between repeats decreased deletion frequency to 65% of that for identical repeats, and frequency was further reduced to about 5% of the identical repeat value when repeats differed by 18%. These findings suggest that HIV-1 recombination is less sensitive to genetic differences than is the wild-type cellular DNA recombination machinery[100,115].

Differences of 27% and greater did not yield detectable HIV-1 recombinants, suggesting repeat deletion was reduced more than 300-fold. However, lower levels of overall sequence similarity may be sufficient to direct detectable homologous recombination *in vivo*, provided the recombinant possesses a selective advantage. Products of recombination between significantly different subtypes[90], and even between M and O group members[82,108], have been reported.

Such recombinants may arise essentially by chance, may rely on residual homology or homology independent recombination triggers, or may be disproportionately common in short regions of high donor/acceptor template sequence identity. It is important to consider that, in the homeologus deletion assay described above, sequence variation was evenly distributed, whereas naturally arising variation can be more clustered. In gammaretrovirus systems, as little as 12 nucleotides of identity between recombination donor and acceptor templates appears sufficient to accurately target homologous recombination, albeit at a lower frequency than that observed for more extensive regions of homology[23,86]. Small patches of sequence identity are sometimes observed at otherwise non-homologous retroviral recombination junctions, suggesting that at least some non-homologous recombination is driven by highly localized regions of homology[128]. Template features besides sequence homology, such as RNA secondary structure, spacing of sequences, homopolymeric stretches, and other less-well defined features of sequence context, appear to modulate recombination frequency[24,59,73,81,131]. The extents to which conserved regions of microhomology, homology-independent recombination hot spots, or other factors that affect biochemical aspects of recombination, contribute to the formation of HIV-1 recombinants will remain unclear until more experimentally derived recombinants have been characterized. Similarly, determining whether or not common crossover junctions observed among clinical isolates are recombination hot spots, or if they are detected solely due to selective advantage, will require experiments designed to test recombination frequencies within the implicated regions.

One implication of the suppressive effects of sequence variation is that reported HIV-1 recombination frequencies (two to three events per replication cycle[42,124] may be underestimates, because they were based on studies with HIV-1 strains that differed by around 5%. Another implication is that the complex mosaic structures often observed for inter-subtype recombinants likely did not arise during a single replication cycle, but instead by sequential co-infection by viruses containing subsets of the observed recombination junctions. However, significant deviation from average crossover frequencies has been observed, even when small numbers of recombinants were studied, with some products displaying at least 10 crossovers even though the average number was less than three[42]. Thus, the range of crossover numbers within recombinant populations is likely to be broad, and multiple crossovers may occasionally occur in single cycles even between two highly divergent strains.

Transduction assays

Transduction assays provide another one-vector approach for studying template switching. These are based on one of the prevailing models for host oncogene capture by simple retroviruses, which postulates that host gene incorporation occurs via non-homologous recombination between viral and host RNAs (Fig. 7a)[128]. This model envisions that host sequences became encapsidated as a 3' appendage to a viral RNA. Retroviral polyadenylation signals are "leaky"[37], and if polyadenylation read-through occurred from a provirus that had integrated upstream of an oncogene, oncogene RNA would be covalently linked to the viral sequences and ferried into viral particles[105]. An additional rare event —non-homologous template switching between sequences in the body of the retrovirus (between viral LTRs) and host sequences in the extended 3' tail— would be required to generate an oncogenic retrovirus with a prototypical acute transforming retrovirus structure[103]. Note that it is not clear whether or not this model was the mechanism involved in the generation of naturally-arising transductants, and that other models for transduction, including ones that envision juxtapositioning of host and viral sequences by RNA splicing or by DNA-level recombination, have also been proposed[30,109]. However, these alternate models do not directly pertain to the assays described here.

Transduction-type homologous recombination vectors incorporate elements of postulated transduction intermediate read-through RNAs, but in-

clude modifications that increase recombination between sequences in the vector body and the appended 3' end (Fig. 7b)[77]. In these vectors, the native polyadenylation signal was inactivated so that all RNAs include the 3' extension. The second modification was to make vector body and the 3' extension sequences homologous to one another. The vector body contained a defective *lac*Z allele, and the patch repair donor in the 3' extension contained *lac*Z sequences that flank the region deleted from the vector body. Recombination, as scored by patch repair of *lacZ*, is surprisingly frequent using these vectors; more than 25% of the puromycin resistant products stain blue with X-gal[77]. Because homologous recombination that did not restore *lac*Z might proceed undetected, in follow-up experiments periodic point mutations were introduced into the patch-repair donor to score crossovers in alternate intervals (An and Telesnitsky, *in preparation*). The introduced mutations decreased blue colony frequency slightly, as homeologus recombination assays would predict[1]. However, analysis of unstained colonies confirmed that these included recombinants at a frequency roughly equal to those detected by blue-white screening. These findings suggest that synthesis of more than half of all parental vector products involved non-selected template switching to 3' appendage sequences that were co-packaged with "genome" (vector body) sequences, but which are not required for provirus formation according to the model presented in figure 2. Transduction assays for recombination offer the advantage of requiring only a single vector. On first consideration, they appear to suffer from some of the same limitations as repeat deletion vectors, since the assayed template switches appear to occur between two positions on a single RNA. However, RT is such a sluggish enzyme that DNA synthesis takes several hours to complete[74]. Each virion contains a large molar excess of RT, and RT's RNaseH and polymerase activities can act sequentially. This provides RNaseH time to degrade template RNA after the first strong-stop template switch, and thus sequences in the 3' appendage and the recombination target are almost certainly on separate RNAs at the time of the assayed switches.

Factors that limit recombination: superinfection and cell co-infection

A fundamental prerequisite for recombination between two RNAs is co-infection on the cellular level. For the two vector assays described here, RNA co-expression is often achieved by unnatural means such as pseudotyping virions to permit infection of non-susceptible cells, or by experimental transfection[123].

In the case of naturally arising intersubtype and some intrasubtype recombinants, superinfection of a patient is an obvious prerequisite to cellular co-infection. The prevalence of recombinants and observations that serial infection can occur, even in the presence of robust immune responses, provide evidence that patient superinfection occurs more readily than had previously been speculated[8,90,91,96]. However, co-infection of an individual does not invariably lead to recombination. For example, no HIV-1/HIV-2 recombinants have been reported, even though dual infections are not rare[66]. The reasons for this are not completely understood, but interference at the level of cell co-infection and differences in RNA packaging are among potential blocks to recombination that have been described[50,90]. The following sections introduce these and other parameters that can limit recombination.

Cell co-infection for simple retroviruses is frequently limited by viral interference: the phenomenon whereby cell infection by one virus limits repeated infection by a second virus that shares the same receptor[116]. Even though HIV-1 is known to down-regulate the surface expression of its receptor and viral interference has been observed *in vitro*, the extent to which cellular resistance to superinfection affects HIV-1 spread *in vivo* is unclear. Some reports suggest that cell co-infection *in vivo* is common[32,47,90].

Recombination between variants that arise within an individual can be detectable without superinfection of a patient, provided cell co-infection occurs, and it is likely that the extent of this has been under-appreciated because the similarity of parental strains masks recombinants[34]. It has been argued that one competitive advantage of recombination is its potential to repair deleterious mutations that arise during error-prone replication, or to promote phenotypic reversion[21]. Such phenomena clearly function during HIV replication *in vivo*, as studies of the dynamics of drug resistance have revealed evidence of intrapatient recombination between circulating strains and archived proviruses[35]. It is interesting to note that as integrated elements, defective proviruses can persist without replication. Some may even have a survival advantage if their defects shield cells that contain them from immune detection. Many patient-derived isolates contain mutations that likely render HIV replication defective[119], but because two different defective retroviruses can readily recombine to form infectious virus, even defective genomes can serve as functional recombination substrates.

RNA co-packaging, heterodimerization, and recombinant detection

Reverse transcription occurs within the viral nucleoprotein complex that forms upon virion entry, and not free in the cytoplasm[9,27]. This is probably the reason genetic recombination occurs during heterozygous virion infection but not if a single cell is co-infected by two different viruses. Although anecdotal evidence for recombination-mediated incorporation of target cell sequences has been reported for alpharetroviruses[75], a gen-

eral requirement for retroviral recombination appears to be virion co-packaging of more than one type of RNA from the producer cell. Thus, one factor that limits the ability of two co-expressed sequences to recombine is whether or not they can become co-packaged in an individual virion.

The genomic RNAs in retroviral particles are packaged as non-covalently linked dimers[92]. Despite significant effort, the nature of the HIV-1 RNA dimer linkage structure is not fully understood. Dimerization is likely nucleated by inter-RNA base-pairing between palindromic loops that cap hairpin structures present on each RNA, followed by an expansion of the dimer interface to form an extended structure (Fig. 8)[33]. The dimer-initiating hairpin, or dimerization initiation site (DIS), is part of the genetically-defined Ψ packaging signal near the 5' end of HIV that is necessary for preferential encapsidation of Ψ+ RNAs (Fig. 8)[7,57]. Dimer linkage and packaging signals are difficult to separate genetically, and are not as easily assigned to a limited genomic region for HIV-1 as for gammaretroviruses like murine leukemia virus (MLV). HIV RNAs that lack Ψ are packaged fairly well and preferentially over bulk mRNA in the absence of Ψ+ RNAs[7,36]. Retroviruses engineered to lack genomic RNAs are not "empty" –even for wild-type viruses, about half the encapsidated RNA in a virion population is non-viral[7].

Although viral RNAs are packaged as dimers, an ability of RNAs to heterodimerize is not absolutely required for co-packaging. Simple retroviruses have provided several examples of recombinogenic patch repair by endogenous retroelements known to be incorporated into retroviral particles but which are unlikely to heterodimerize with virus genomes, and RNAs, which presumably only occasionally and by chance become encapsidated, can also recombine with viral genomes[67]. Co-packaged RNAs can recombine and form a functional provirus if at least one of the RNAs contains *cis*-acting sequences necessary for replication. It has been reported that encapsidated RNAs that lack retroviral *cis*-acting signals can be converted into DNA and integrated like proviruses in certain circumstances[60].

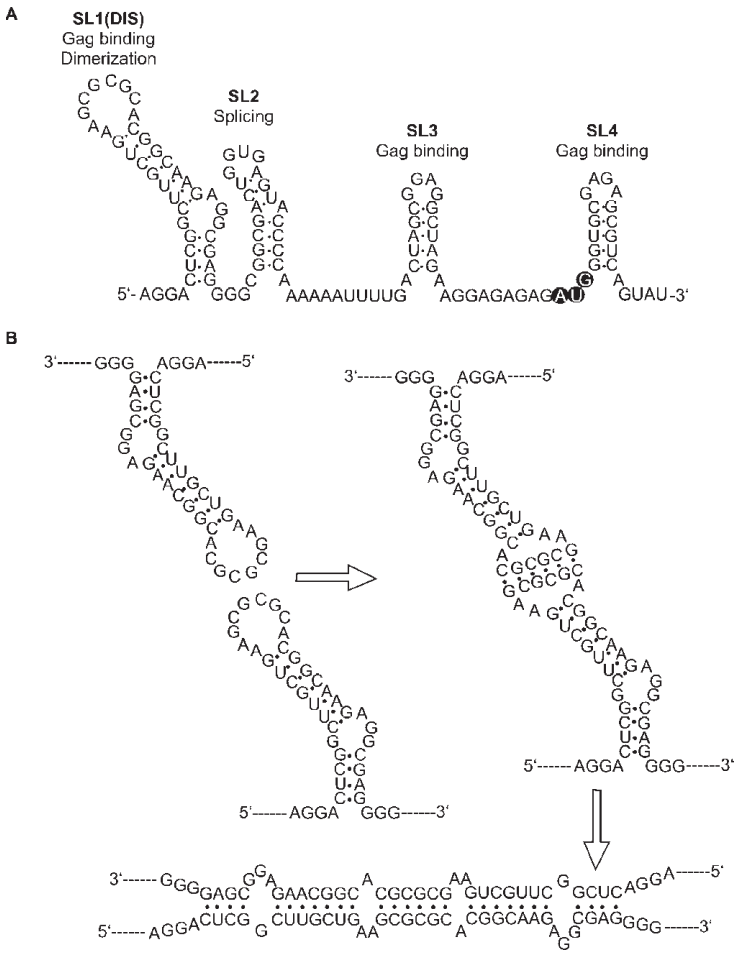

Figure 8. *RNA structures, including the dimerization initiation sites, in the dimerization/packaging region near the 5' end of HIV-1 RNAs. A) Putative stem-loop structures of sequences in the vicinity of the dimerization initiation site (DIS, also known as stem-loop 1 or SL1) of subtype B HIV-1 genomic RNA, and their implicated functions. B) Schematic representation of the "kissing" interaction between the DIS palindromic loops of two RNAs and postulated isomerized structure. Adapted from[17].*

Although recombinants between RNAs that cannot dimerize have been observed, it has been assumed that these products arise only rarely. An untested assumption used in recombination frequency calculations has been that when two retroviral RNAs are co-packaged, they are generally packaged as partners in heterodimers[19]. Because recombination is only detectable if more than one kind of RNA is co-packaged into a virion, this assumption predicts that if two RNAs cannot heterodimerize, recombination between them will be vastly reduced.

St. Louis, et al. approached the question of RNA heterodimerization effects on recombination rates by studying HIV-1 recombination in tissue culture with RNAs whose dimer initiation signals were believed incapable of heterodimerization[102]. Among natural HIV-1 isolates, the palindromic sequences that cap DIS stem-loops vary so that, for example, a subtype B RNA should be incapable of dimerizing with a subtype A RNA. If such sequence differences do prevent heterodimerization, and if heterodimerization is a normal prerequisite to recombination, recombination frequencies between RNAs that differ in palindromic sequences should be much lower than for RNAs with the same sequences. Because clinical isolates resulting from recombination between seemingly heterodimer-resistant strains were known, it was clear that such differences did not block recombination entirely. The St. Louis, et al. studies sought to address if and/or how much effect DIS variation had on recombination frequencies. In these studies, restriction site polymorphisms were engineered into pairs of HIV vectors that either possessed the same DIS or that possessed different subtype specific DIS regions, and restriction site co-segregation patterns were examined to monitor recombination. Although experiments were performed with replication-competent virus, which limited measurement precision, the results clearly demonstrated that the reduction in recombination frequency associated with differing DIS was much lower than predicted if DIS palindrome-mediated heterodimerization were essential[102].

The biology behind these findings remains unclear. Work with DIS-containing transcripts in purified *in vitro* reactions has demonstrated the importance of the DIS to dimerization[33], so it seemed reasonable to assume that differing DIS would inhibit recombination. Whether or not subtype-specific differences in DIS loops are sufficient to interfere with heterodimerization during viral replication remains to be established. If they do, the findings of St. Louis, et al. suggest that HIV-1 RNAs that cannot heterodimerize readily can, nonetheless, become co-packaged[102].

Differences between HIV-1 and simple retrovirus recombination

Despite the complex mosaic structures observed in many natural HIV recombinants, the high frequency of HIV-1 genetic recombination measured during single replication cycles initially was surprising[124]. This was because it was already well established that gammaretrovirus recombination was far less frequent[41] and it had been assumed that HIV-1 frequencies would be similar. Virus species-specific differences in template switching properties initially appeared the most likely cause of recombination-rate differences between HIV-1 and gammaretroviruses[42]. HIV-1 reverse transcription is known to differ from that of gammaretroviruses in such ways as the number and specificity of its primers, and several reports have suggested mechanistic differences between HIV-1 and simple retrovirus RTs[109]. Accessory factors such as Nef and Vpr, which are absent from simple retroviruses, appear to affect HIV-1 reverse transcription and thus might affect recombination[63,99]. However, ongoing work from our laboratory now suggests that such enzymological differences are not the primary cause of genetic recombination-rate differences[77].

Our conclusion that HIV-1 and MLV differ in recombination rates, but perform template switches at similar frequencies, is based on a series of single replication cycle assays. We first developed repeat-deletion assays to compare HIV-1 and MLV. Because it had been demonstrated that HIV-1 recombines more frequently than MLV[4,124], we had anticipated observing a higher repeat-deletion rate for HIV-1 than for MLV. However, in contrast to expectations, repeat-deletion frequencies for the two viruses were very similar[2]. We next established two-vector recombination assays for both HIV-1 and MLV, to address whether or not reported intermolecular recombination differences were reproducible[77]. After we observed intermolecular recombination rate differences consistent with previous reports, we compared template-switching properties using the transduction-type assay vectors described above. Template-switching frequencies for HIV-1 and MLV were indistinguishable using transduction vectors[77]. This suggested that whereas recombination rates, as measured by co-segregation of markers from two different RNAs, differed dramatically for HIV and MLV, intrinsic template-switching properties for the two virus species were very similar.

One possible explanation for why HIV and MLV recombination rates might differ, even though their template-switching frequencies are similar, is based on reported genetic findings about gammaretrovirus recombination. Calculated overall genetic recombination frequencies for gammaretroviruses are relatively low –about one crossover per three proviruses– but the distribution of observed genetic crossovers is highly non-random[5,21,39]. The vast majority of gammaretroviruses show no genetic evidence of recombination. However, among recombinants, most display evidence of more than one crossover, and some display several. These genetic observations have been argued to support a "high negative interference" model for retroviral recombination, which suggests that genetic recombination is not a routine process but instead bipha-

sic, with crossovers during reverse transcription either frequent or else rare[21]. It has been proposed that the performance of one recombination event predisposes the ordinarily recombination-proof virion to performing additional switches, or alternatively that only a subset of virions are prone to generating recombinant products[5,39]. It has further been suggested that the viral nucleoprotein complex architecture of one subset of particles is somehow aberrant, which results in a high frequency of recombination among products of virions with this postulated defect[44].

An alternate hypothesis for why only a subset of gammaretroviruses display evidence of recombination, is that only a subset of particles co-packages two different RNAs[19]. Retroviral RNA trafficking —including which host factors are involved, which subset of RNAs are packagable, the intracellular compartment where commitment to packaging is made, and whether or not *cis*-acting packaging sequences are sufficient for RNA packaging— differs dramatically among retroviral species[6,12,22]. As mentioned above, recombinogenic template switching between co-packaged RNAs is only detectable if the co-packaged RNAs differ. If two different RNAs become co-packaged less frequently than random co-packaging would predict, the following alternate hypothesis for gammaretrovirus recombination emerges. Recombinogenic template switching occurs at a uniform and high distribution of frequencies for all virions. However, recombination is detectable only for the small subset of virions that co-package two different RNAs.

Ongoing experiments in our laboratory have used biochemical approaches to assess the composition of virion RNA dimers (Flynn and Telesnitsky, *in preparation*). The results suggest that when two different MLV Ψ+ RNAs are co-expressed, they generally become encapsidated as RNA homodimers. In contrast, HIV-1 co-expressed RNAs are more likely to heterodimerize, possibly indicating that RNA commitment to dimerization occurs at different replication stages for these two viruses. We postulate that these differences in randomness of co-expressed RNA dimer partner selection contribute to differences in genetic recombination rates. The implication that MLV preferentially co-packages two identical RNAs, but HIV-1 co-packages co-expressed RNAs more randomly, also provides a plausible explanation for reported differences in minus strong-stop template switching. Genetics suggest this replicative switch occurs essentially only between the two ends of a single RNA for gammaretroviruses, but can occur either between the two ends of one RNA, or from one to another co-packaged RNA in the case of HIV-1[110,113,124]. Our new results suggest that this observed genetic difference may reflect differing virion RNA composition rather than differences in template switching. If the majority of gammaretrovirus particles are homozygous, template switching between two co-packaged RNAs would be indistinguishable genetically from switching between the two ends of a single RNA.

As introduced above, the RNA content of retroviral particles is not tightly fixed. It remains unclear whether or not most retroviruses contain precisely two RNAs, or if variation –from zero to four or more co-packaged RNAs– is normal. Both lentiviruses and gammaretroviruses are capable of packaging Ψ+ RNAs more than twice the native length[54,101], and alpharetroviruses may co-package more than two RNAs when RNAs are short[97]. Findings that host factors can modulate the molar ratio of RNA to virion protein in HIV-1 particles support the notion that lentiviruses share the property of flexibility in packaged RNA composition[69]. Our two-vector recombination assays suggest that MLV recombination is about 6-fold less frequent than that of HIV-1, but our ongoing dimer detection assays suggest that MLV RNA heterodimer formation may be reduced more than 6-fold relative to predicted random RNA dimer formation ratios[77]. This suggests the possibility that, in at least some instances, the recombination substrates for MLV may be two co-packaged homodimers rather than two different RNAs co-packaged in a single heterodimer.

If HIV-1 can co-package more than one dimer, speculatively, it may be possible for HIV-1 to co-package –and to recombine in a single cycle– more than two genetically distinct RNAs. It has been suggested that the appearance of sequences from more than two parental strains in many HIV-1 recombinants likely reflects recombination between pre-existing recombinants rather than triple or higher multiplicities of infection[90]. However, recent *in situ* hybridization experiments with patient-derived single cells suggest it may not be unusual for individual cells to contain four or more different proviruses *in vivo*[32,47]. Taken together, these observations leave open the possibility that recombination between more than two parental strains may sometimes occur in a single round of reverse transcription.

Implications and limitations of experimental recombination systems

An obvious limitation of the recombination assays presented above is that they describe HIV-1 replication in cultured cells. Single-cycle experiments are intended to be largely free of selective pressures. Because small differences in selective advantage can have large effects on replicating populations[20], tissue culture-based results obviously do not predict outcomes of natural infection. What these experiments do provide is baseline rates of recombination that can be compared to clinical outcomes to help assess effects of immunologic and other selective forces *in vivo*.

These assays are also limited on the opposite end of the spectrum, in that intact cells contain too many unknown components to allow precise probing of mechanistic details of template switching. Many studies designed to examine RT prop-

erties related to genetic recombination have been performed with purified enzymes and model primer-templates in reconstituted reactions *in vitro,* and have recently been reviewed elsewhere[72,73]. The findings from such systems vary in the extents to which they recapitulate events as they occur during viral replication, but they provide valuable insight into mechanistic features of template switching. The type of cell a virus infects can affect reverse transcription outcomes. Single replication cycle assays have been used to address the effects of altering the intracellular environment on recombination frequencies. The immortalized tissue culture cells used in most experiments described above are metabolically highly active. Compared to tissue culture cells, HIV-1 reverse transcription takes longer and/or is disrupted in less active cells[88,125]. In gammaretrovirus recombination experiments, where the time required to complete reverse transcription was tripled by serum starvation or by treating cells with hydroxyurea, the frequency of template switching increased 3-fold over frequencies observed in untreated cells[87,104]. Treatment of HIV-1 infected tissue culture cells with either hydroxyurea or AZT also modestly, but reproducibly, stimulated HIV-1 template switching[2]. These findings suggest that frequencies of HIV-1 genetic recombination may differ significantly from one cellular context or host compartment to another during natural infection. Because HIV-1 DNA synthesis may take up to ten times as long to complete in resting cells than in actively dividing ones[88], intra-host conditions may increase the frequency of recombination even more than the 3-fold achieved by experimental substrate limitation[87,104]. The numbers of crossovers per genome observed among 86 tested HIV-1 recombinants generated in immortalized cells ranged from one to ten, demonstrating that at least 10 crossovers can occur per HIV replication cycle in rapidly dividing cells[42]. This observed range of experimental crossover frequencies, paired with observations of increased recombination rates when reverse transcription is slowed, makes it reasonable to postulate that among the recombinants generated in multiply infected patients, there may be some whose synthesis involved 30 or more crossovers during a single replication cycle.

Conclusions

The variability of retroviral quasispecies has long been recognized. One factor that contributes to this genetic diversity is high base substitution rates: an average of 0.1 to one substitution per genome per generation[89]. The ability of HIV-1 to undergo genetic recombination has also been recognized, but it has until recently remained unclear whether recombination was frequent, or if it were a relatively rare occurrence that was detectable only because of selective advantage gained by genome rearrangement. The experimental work summarized above suggests that some recombinants —those between significantly divergent genomes or between sequences that are not ordinarily co-packaged into individual HIV-1 virions— are generated only rarely. When such recombinants are detected in viral populations, this likely reflects both selective pressures and the exceptional level of replication that is characteristic of HIV-1. In contrast, these experimental findings suggest that recombination between many co-circulating strains occurs at a remarkably high frequency, with an average of more than three crossover events during the synthesis of every provirus formed in a rapidly-dividing cell, and even more frequently in a less metabolically-active cell. These findings suggest that genetic recombination among similar strains, such as the quasispecies that develop longitudinally within individual patients, likely occurs whenever cell co-infection provides an opportunity.

Acknowledgments

We thank Denise Kirschner for useful discussion of mathematical interpretations of data, Joseph Dougherty for sharing results prior to publication, and David Friedman for critical reading of the manuscript. We apologize to those whose work has been fundamental to virus biology but is acknowledged only indirectly by reference to review articles. Supported by the Great Lakes Center for AIDS Research and by GM64479 from the National Institutes of Health.

References

1. An W, Telesnitsky A. The effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a HIV type 1 vector. J Virol 2002;76:7897-902.
2. An W, Telesnitsky A. Frequency of direct repeat deletion in a HIV type I vector during reverse transcription in human cells. Virology 2001;286:475-82.
3. Anderson J, Bowman E, Hu W. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation. J Virol 1998;72:1195-202.
4. Anderson J, Pathak V, Hu W. Effect of the murine leukemia virus extended packaging signal on the rates and locations of retroviral recombination. J Virol 2000;74:6953-63.
5. Anderson J, Teufel R, Yin P, Hu W. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. J Virol 1998;72:1186-94.
6. Banks J, Kealoha B, Linial M. An MΨ-containing heterologous RNA, but not *env* mRNA, is efficiently packaged into avian retroviral particles. J Virol 1999;73:8926-33.
7. Berkowitz R, Fisher J, Goff S. RNA packaging. Curr Top Microbiol Immunol 1996;177-218.
8. Blackard J, Cohen D, Mayer K. HIV superinfection and recombination: current state of knowledge and potential clinical consequences. Clin Infect Dis 2002;34:1108-14.
9. Bowerman B, Brown P, Bishop J, Varmus H. A nucleoprotein complex mediates the integration of retroviral DNA. Genes and Development 1989;3:469-78.

10. Brincat J, Pfeiffer J, Telesnitsky A. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication. J Virol 2002;76:88-95.
11. Buiser R, DeStefano J, Mallaber L, Fay P, Bambara R. Requirements for the catalysis of strand transfer synthesis by retroviral DNA 11. polymerases. J Biol Chem 1991;266:13103-9.
12. Butsch M, Boris-Lawrie K. Destiny of unspliced retroviral RNA: Ribosome and/or virion? J Virol 2002;76:3089-94.
13. Champoux J. Roles of ribonuclease H in reverse transcription. In: Skalka AM, Goff SP (ed). Reverse transcriptase. Cold Spring Harbor: Cold Spring Harbor Laboratory 1993:103-117.
14. Charneau P, Alizon M, Clavel F. A second origin of DNA plus-strand synthesis is required for optimal HIV replication. J Virol 1992;66:2814-20.
15. Charneau P, Clavel F. A single-stranded gap in HIV unintegrated linear DNA defined by a central copy of the polypurine tract. J Virol 1991;65:2415-21.
16. Clavel F, Hoggan M, Willey R, Strebel K, Martin M, Repaske R. Genetic recombination of HIV. J Virol 1989;63:1455-9.
17. Clever J, Parslow T. Mutant HIV type 1 genomes with defects in RNA dimerization or encapsidation. J Virol 1997; 71:3407-14.
18. Coffin J. Retroviridae: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM, et al. (ed). Fundamental Virology. 3rd ed. Philadelphia: Lippincott.
19. Coffin J. Structure of retroviral genomes. In: Weiss R, Teich N, Varmus H, Coffin J (ed). RNA Tumor Viruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratories 1982:261-368.
20. Coffin J. HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis and therapy. Science 1995;267:483-9.
21. Coffin J. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J Gen Virol 1979; 42:1-26.
22. Cullen B. Nuclear RNA export pathways. Molecular and Cellular Biology 2000;20:4181-7.
23. Dang Q, Hu W. Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. J Virol 2001;75:809-20.
24. Delviks K, Pathak V. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. J Virol 1999; 73:7923-32.
25. DeStefano J, Bambara R, Fay P. The mechanism of HIV reverse-transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates. J Biol Chem 1994;269:161-8.
26. DeStefano J, Raja A, Cristofaro J. *In vitro* strand transfer from broken RNAs results in mismatch but not frameshift mutations. Virology 2000;276:7-15.
27. Fassati A, Goff S. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. J Virol 2001;75:3626-35.
28. Garcia-Lerma J, Heneine W. Resistance of HIV type 1 to reverse transcriptase and protease inhibitors: genotypic and phenotypic testing. J Clinical Virology 2001;21:197-212.
29. Gilboa E, Mitra S, Goff S, Baltimore D. A detailed model of reverse transcription and tests of crucial aspects. Cell 1979;18:93-100.
30. Goldfarb M, Weinberg R. Generation of novel, biologically active Harvey sarcoma viruses via apparent illegitimate recombination. J Virol 1981;38:136-50.
31. Goldfarb M, Weinberg R. Structure of the provirus within NIH 3T3 cells transfected with Harvey sarcoma virus DNA. J Virol 1981;38:125-35.
32. Gratton S, Cheynier R, Dumaurier M, Oksenhendler E, Wain-Hobson S. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. Proc Natl Acad Sci USA 2000;97:14566-71.
33. Greatorex J, Lever A. Retroviral RNA dimer linkage. J Gen Virol 1998;79:2877-82.
34. Groenink M, Andeweg A, Fouchier R, et al. Phenotype-associated env gene variation among eight related HIV type 1 clones: evidence for *in vivo* recombination and determinants of cytotropism outside the V3 domain. J Virol 1992;66:6175-80.
35. Gunthard H, Leigh-Brown A, D'Aquila R, et al. Higher selection pressure from antiretroviral drugs *in vivo* results in increased evolutionary distance in HIV-1. J Virol 1999; 259:154-65.
36. Helga-Maria C, Hammarskjold M, Rekosh D. An intact TAR element and cytoplasmic localization are necessary for efficient packaging of HIV type 1 genomic RNA. J of Virol 1999;73:4127-35.
37. Herman S, Coffin J. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. Science 1987;236:845-8.
38. Hizi A, Tal R, Shaharabany M, Loya S. Catalytic properties of the reverse transcriptases of HIV type 1 and type 2. J Biol Chem 1991;266:6230-9.
39. Hu W, Bowman E, Delviks K, Pathak V. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. J Virol 1997;71:6028-36.
40. Hu W, Temin H. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci USA 1990;87:1556-60.
41. Hu W, Temin H. Retroviral recombination and reverse transcription. Science 1990;250:1227-33.
42. Jetzt A, Yu H, Klarmann G, Ron Y, Preston B Dougherty J. High rate of recombination throughout the HIV type 1 genome. J Virol 2000;74:1234-40.
43. Jones J, Allan R, Temin H. Alteration of location of dimer linkage sequence in retroviral RNA: little effect on replication or homologous recombination. J Virol 1993;67:3151-8.
44. Jones J, Allan R, Temin H. One retroviral RNA is sufficient for synthesis of viral DNA. J Virol 1994;68:207-16.
45. Julias J, Ferris A, Boyer P, Hughes. Replication of phenotypically mixed HIV-1 virions containing catalytically active and catalytically inactive reverse transcriptase. J Virol 2001;75:6537-46.
46. Julias J, Hash D, Pathak V. E-vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy. J Virol 1995;69:6839-46.
47. Jung A, Maier R, Vartanian J, et al. Multiply infected spleen cells in HIV patients. Nature 2002;48:144.
48. Junghans R, Boone L, Skalka A. Retroviral DNA H structures: displacement-assimilation model of recombination. Cell 1982;30:53-62.
49. Katz R, Skalka A. Generation of diversity in retroviruses. 1990;24:409-45.
50. Kaye J, Lever A. Nonreciprocal packaging of HIV type 1 and type 2 RNA: a possible role for the p2 domain of Gag in RNA encapsidation. J Virol 1998;72:5877-84.
51. Kellam P, Larder B. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J Virol 1995;69:669-74.
52. Kinoshita K, Honjo T. Linking class-switch recombination with somatic hypermutation. Nat Rev Mol Cell Biol 2001;2:493-503.
53. Topping K, Telesnitsky A. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. EMBO J 1997;16:856-65.
54. Kumar M, Keller B, Makalou N, Sutton R. Systematic determination of the packaging limit of lentiviral vectors. Human Gene Therapy 2001;12:1893-905.
55. Kung H, Fung Y, Majors J, Bishop J, Varmus H. Synthesis of plus strands of retroviral DNA in cells infected with avian sarcoma virus and mouse mammary tumor virus. J Virol 1981;37:127-38.
56. Lercher M, Hurst L. Human SNP variability and mutation rate are higher in regions of high recombination. Trends in genetics 2002;18:337-40.
57. Lever A, Gottlinger H, Haseltine W, Sodroski J. Identification of a sequence required for efficient packaging of HIV type 1 RNA into virions. J Virol 1989;63:4085-7.
58. Li T, Zhang J. Determination of the frequency of retroviral recombination between two identical sequences within a provirus. J Virol 2000;74:7646-50.

59. Li Y, Carpenter S. *Cis*-acting sequences may contribute to size variation in the surface glycoprotein of bovine immunodeficiency virus. J Gen Virol 2001;82:2989-98.
60. Linial M, Medeiros E, Hayward W. An avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: biological and biochemical characterization. Cell 2001;15:1371-81.
61. Lovett S, Drapkin P, Sutera V, et al. A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in *Escherichia coli*. Genetics 1993;135:631-42.
62. Malim M, Emerman M. HIV-1 sequence variation: drift, shift, and attenuation. Cell 2001;104:469-72.
63. Mansky L. The mutation rate of HIV type 1 is influenced by the *vpr* gene. Virology 1996;222:391-400.
64. McCutchan F. Global Diversity of HIV. In: Crandall KA (ed). The Evolution of HIV. Baltimore: The Johns Hopkins University Press 1999:41-101.
65. McCutchan F. Understanding the genetic diversity of HIV-1. AIDS 2000;14:31-44.
66. McGrath K, Hoffman N, Resch W, Nelson J, Swanstrom. Using HIV-1 sequence variability to explore virus biology. Virus Research 2001;76:137-60.
67. Mikkelsen J, Pedersen F. Genetic reassortment and patch repair by recombination in retroviruses. J Biomed Science 2000;7:77-99.
68. Miller A. Development and applications of retroviral vectors. In: Coffin JM, Hughes SH, Varmus HE (ed). Retroviruses. Plainview, NY: Cold Spring Harbor Laboratory Press 1997:437-73.
69. Mouland A, Mercier J, Luo M, Bernier L, DesGroseillers Cohen E. The double-stranded RNA-binding protein Staufen is incorporated in HIV type 1: evidence for a role in genomic RNA encapsidation. J Virol 2000;74:5441-51.
70. Moutouh L, Corbeil J, Richman D. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. Proc Natl Acad Sci USA 1996;93:6106-11.
71. Mules E, Uzun O, Gabriel A. Replication errors during *in vivo* Ty1 transposition are linked to heterogeneous RNase H cleavage sites. Mol Cell Biol 1998;18:1094-104.
72. Negroni M, Buc H. Mechanisms of retroviral recombination. Annual Review of Genetics 2001;35:275-302.
73. Negroni M, Buc H. Retroviral recombination: what drives the switch? Mol Cel Biol 2001;2:151-5.
74. O'Brien W, Namazi A, Kalhor H, Mao S, Zack J, Chen I. Kinetics of HIV type 1 reverse transcription in blood mononuclear phagocytes are slowed by limitations of nucleotide precursors. J Virol 1994;68:1258-63.
75. Olsen J, Bova-Hill C, Grandgenett D, Quinn T, Manfredi J Swanstrom R. Rearrangements in unintegrated retroviral DNA are complex and are the result of multiple genetic determinants. J Virol 1990;64:5475-84.
76. Olson P, Temin H, Dornberg R. Unusually high frequency of reconstitution of long terminal repeats in U3-minus retrovirus vectors by DNA recombination or gene conversion. J Virol 1992;66:1336-43.
77. Onafuwa A, An W, Robson N, Telesnitsky A. HIV-1 genetic recombination is more frequent than that of Moloney murine leukemia virus despite similar template switching rates (submitted).
78. Ostertag E, Kazazian H. Twin priming: a proposed mechanism of the creation of inversions in L1 retrotransposition. Genome Research 2001;11:2059-65.
79. Panganiban A, Fiore D. Ordered interstrand and intrastrand DNA transfer during reverse transcription. Science 1988; 241:1064-9.
80. Patel P, Preston B. Marked infidelity of HIV type 1 reverse transcriptase at RNA and DNA template ends. Proc Natl Acad Sci USA 1994;91:549-53.
81. Pathak V, Temin H. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: Deletions and deletions with insertions. Proc Natl Acad Sci USA 1990;87:6024-8.
82. Peeters M, Liegeois F, Torimiro N, et al. Characterization of a highly replicative intergroup M/O HIV type 1 recombinant isolated from a Cameroonian patient. J Virol 1999;73:7368-75.
83. Peliska J, Benkovic S. Fidelity of *in vitro* DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Biochemistry 1994;33:3890-5.
84. Peliska J, Benkovic S. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science 1992;258:1112-8.
85. Pfeiffer J, Georgiadis M, Telesnitsky A. Structure-based Moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct repeat deletion frequencies. J Virol 2000;74:9629-36.
86. Pfeiffer J, Telesnitsky A. Effects of limiting homology on intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. J Virol 2001;75:11263-74.
87. Pfeiffer J, Topping R, Shin N, Telesnitsky A. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription. J Virol 1999;73:8441-7.
88. Pierson T, Zhou Y, Kieffer T, Ruff C, Buck C, Siciliano R. Molecular characterization of preintegration latency in HIV type 1 infection. J Virol 2002;76:8518-31.
89. Preston B, Dougherty J. Mechanisms of retroviral mutation. Trends in Microbiology 1996;4:16-21.
90. Quiñones-Mateu M, Arts E. Recombination in HIV-1: update and implications. AIDS Rev 1999;1:89-100.
91. Ramos A, Hu D, Nguyen L, et al. Intersubtype HIV type 1 superinfection following seroconversion to primary infection in two injection drug users. J Virol 2002;76:7444-52.
92. Rein A. Retroviral RNA packaging: a review. Arch Virol (Suppl) 1994;9:513-22.
93. Rein A, Henderson L, Levin J. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. TIBS 1998;23:297-301.
94. Rhode B, Emerman M, Temin H. Instability of large direct repeats in retrovirus vectors. J Virol 1987;61:925-7.
95. Ricchetti M, Buc H. A reiterative mode of DNA synthesis adopted by HIV-1 reverse transcriptase after a misincorporation. Biochemistry 1996;35:14970-83.
96. Robertson D, Gao F, Hahn B, Sharp P. Intersubtype recombinant HIV-1 sequences. In: Korber B, et al. (ed). Human retroviruses and AIDS 1997: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory 1997.
97. Sakalian M, Wills J, Vogt V. Efficiency and selectivity of RNA packaging by Rous sarcoma virus gag deletion mutants. J Virol 1994;68:5969-81.
98. Schatz O, Cromme R, Naas T, Lindemann D, Mous J, Le Grice S. Inactivation of the RNaseH domain of HIV-1 reverse transcriptase blocks viral infectivity. In: Papas T (ed). Oncogenesis and AIDS. Texas: Portfolio Publishing 1990.
99. Schwartz O, Marechal V, Danos O. HIV type 1 *nef* increases the efficiency of reverse transcription in the infected cell. J Virol 1995;69:4053-9.
100. Shen P, Huang H. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 1986;112:441-57.
101. Shin N, Hartigan-O'Connor D, Pfeiffer J. Replication of lengthened Moloney murine leukemia virus genomes is impaired at multiple stages. J Virol 2000;74:2694-702.
102. St. Louis D, Gotte D, Sanders-Buell D. Infectious molecular clones with the nonhomologous dimer initiation sequences found in different subtypes of HIV type 1 can recombine and initiate a spreading infection *in vitro*. J Virol 1998; 72:3991-8.
103. Sugden B. How some retroviruses got their oncogenes. TIBS 1993;18:233-5.
104. Svarovskaia E, Delviks K, Hwang C. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. J Virol 2000;74:7171-8.
105. Swain A, Coffin J. Mechanism of transduction by retroviruses. Science 1992;255:841-5.
106. Swanstrom R, Bishop J, Varmus H. Structure of a replication intermediate in the synthesis of Rous Sarcoma virus DNA *in vivo*. 1982;42:337-41.
107. Takehisa J, Osei-Kwasi M, Ayisi N, et al. Phylogenetic analysis of HIV type 2 in Ghana and intrasubtype recombination in HIV type 2. AIDS Res Hum Retroviruses 1997;13:621-3.

108. Takehisa J, Zekeng L, Ido E, et al. HIV type 1 intergroup (M/O) recombination in Cameroon. J Virol 1999;73:6810-20.
109. Telesnitsky A, Goff S. Reverse transcriptase and the generation of retroviral DNA. In: Coffin J, Hughes S, Varmus H (ed). Retroviruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory 1997:121-60.
110. Telesnitsky A, Goff S. Strong-stop strand transfer during reverse transcription. In Skalka AM, Goff SP (ed). Reverse transcriptase. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1993:49-83.
111. Temin H. Retrovirus variation and reverse transcription: Abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci USA 1993;90:6900-3.
112. Tisdale M, Schulze T, Larder B, Moelling K. Mutations within the RNase H domain of HIV type 1 reverse transcriptase abolish virus infectivity. J Gen Virol 1991;72:59-66.
113. Van Wamel J, Berkhout B. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. Virology 1998;244:245-51.
114. Vogt V. Retroviral virions and genomes. In: Coffin JM, Hughes SH, Varmus HE (ed). Retroviruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory 1997:27-69.
115. Waldman A, Liskay R. Differential effects of base-pair mismatch on intrachromosoma versus extrachromosomal recombination in mouse cells. Proc Natl Acad Sci USA 1987;84:5340-4.
116. Weiss R. Experimental biology and assay of RNA tumor viruses. In: Weiss R, Teich N, Varmus H, Coffin J. (ed). RNA Tumor Viruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratories 1982:209-60.
117. Wooley D, Bircher L, Smith R. Retroviral recombination is nonrandom and sequence dependent. Virology 1998;243:229-34.
118. Wooley D, Smith R, Czajak S, Desrosiers R. Direct demonstration of retroviral recombination in a rhesus monkey. J Virol 1997;71:9650-3.
119. Wrobel J, Chao S, Conrad M, et al. A genetic approach for identifying critical residues in the fingers and palm subdomains of HIV-1 reverse transcriptase. Proc Natl Acad Sci USA 1998;95:638-45.
120. Wyke J, Beamand J. Genetic recombination in Rous sarcoma virus: the genesis of recombinants and lack of evidence for linkage between pol, env and src genes in three factor crosses. J Gen Virol 1979;43:349-64.
121. Wyke J, Bell J, Beamand J. Genetic recombination among temperature-sensitive mutants of Rous sarcoma virus. Cold Spring Harbor Symposium on Quantitative Biology 1974; 39:897-905.
122. Xu H, Boeke J. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 1987; 84:8553-7.
123. Yu H, Jetzt A, Dougherty J. Use of single-cycle analysis to study rates and mechanisms of retroviral mutation. Methods 1997;12:325-36.
124. Yu H, Jetzt A, Ron Y, Preston B, Dougherty J. The nature of HIV type 1 strand transfers. J Biol Chem 1998;273:28384-91.
125. Zack J, Arrigo S, Weitsman S, Go A, Haislip A, Chen I. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell 1990;61:213-22.
126. Zhang J, Sapp C. Recombination between two identical sequences within the same retroviral RNA molecule. J Virol 1999;73:5912-7.
127. Zhang J, Tang LY, Li T, Ma Y, Sapp C. Most retroviral recombinations occur during minus-strand DNA synthesis. J Virol 2000;74:2313-22.
128. Zhang J, Temin H. 3' junctions of oncogene-virus sequences and the mechanisms for formation of highly oncogenic retroviruses. J Virol 1993;67:1747-51.
129. Zhang J, Temin H. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science 1993;259:234-8.
130. Zhang J, Temin H. Retrovirus recombination depends on the length of sequence identity and is not error prone. J Virol 1994;68:2409-14.
131. Zhang W, Hwang C, Hu W, Gorelick R, Pathak V. Zinc finger domain of murine leukemia virus nucleocapsid protein enhances the rate of viral DNA synthesis *in vivo*. J Virol 2002;76:7473-84.
132. Zhang W, Svarovskaia E, Barr R, Pathak V. Y586F mutation in murine leukemia virus reverse transcriptase decreases fidelity of DNA synthesis in regions associated with adenine-thymine tracts. PNAS 2002;99:10090-5.
133. Zhuang J, Jetzt A, Sun G, et al. Human immunodeficiency virus type-1 recombination: rate, fidelity, and putative hotspots. J Virol 2002;76:11273-82.