Influence of Host Gene Transcription Level and Orientation on HIV-1 Latency in a Primary-Cell Model[∇]

Liang Shan,^{1,2} Hung-Chih Yang,^{1†} S. Alireza Rabi,¹ Hector C. Bravo,³ Neeta S. Shroff,⁵
Rafael A. Irizarry,³ Hao Zhang,⁴ Joseph B. Margolick,⁴
Janet D. Siliciano,¹ and Robert F. Siliciano^{1,5*}

Departments of Medicine¹ and Department of Pharmacology and Molecular Sciences,² Johns Hopkins University School of Medicine, Department of Biostatistics³ and Department of Molecular Microbiology and Immunology,⁴ Johns Hopkins University Bloomberg School of Public Health, and Howard Hughes Medical Institute,⁵ Baltimore, Maryland 21205

Received 5 December 2010/Accepted 11 March 2011

Human immunodeficiency virus type 1 (HIV-1) establishes a latent reservoir in resting memory CD4⁺ T cells. This latent reservoir is a major barrier to the eradication of HIV-1 in infected individuals and is not affected by highly active antiretroviral therapy (HAART). Reactivation of latent HIV-1 is a possible strategy for elimination of this reservoir. The mechanisms with which latency is maintained are unclear. In the analysis of the regulation of HIV-1 gene expression, it is important to consider the nature of HIV-1 integration sites. In this study, we analyzed the integration and transcription of latent HIV-1 in a primary CD4⁺ T cell model of latency. The majority of integration sites in latently infected cells were in introns of transcription units. Serial analysis of gene expression (SAGE) demonstrated that more than 90% of those host genes harboring a latent integrated provirus were transcriptionally active, mostly at high levels. For latently infected cells, we observed a modest preference for integration in the same transcriptional orientation as the host gene (63.8% versus 36.2%). In contrast, this orientation preference was not observed in acutely infected or persistently infected cells. These results suggest that transcriptional interference may be one of the important factors in the establishment and maintenance of HIV-1 latency. Our findings suggest that disrupting the negative control of HIV-1 transcription by upstream host promoters could facilitate the reactivation of latent HIV-1 in some resting CD4⁺ T cells.

Human immunodeficiency virus type 1 (HIV-1) establishes a stable latent reservoir in resting memory CD4⁺ T cells that persists in patients on highly active antiretroviral therapy (HAART) and that is able to produce replication-competent virus following cellular activation (7–9, 14, 61). The latent reservoir is a major barrier to virus eradication (15, 49, 50). In this stable reservoir, the provirus is transcriptionally silent (10, 22). This latent reservoir may contribute to the rapid rebound of viremia after cessation of HAART (26, 63). Deliberately inducing the reactivation of latent HIV-1 may be required to effectively target the reservoir and achieve eradication. Understanding the mechanisms that maintain HIV-1 latency is critical to this therapeutic strategy.

Mechanisms that maintain HIV-1 latency *in vivo* are incompletely understood. It is widely accepted that the lack of active forms of key cellular transcription factors (3, 13, 16, 31, 42, 58) and of the HIV-1 Tat protein and its cellular cofactors (12, 23, 25, 29, 47, 51) limits the initiation and elongation, respectively, of viral transcription in resting CD4⁺ T cells (32, 60). In addition, DNA methylation and repressive histone modifications, especially the formation of a single nucleosome located at the

viral promoter, have been postulated to promote transcriptional silencing of integrated proviruses (2, 11, 21, 30, 52, 55, 59). Posttranscriptional mechanisms may also play a role (24, 33, 43).

In the analysis of the regulation of HIV-1 gene expression, it is important to consider the nature of HIV-1 integration sites. Pioneering studies by Bushman and colleagues demonstrated that HIV-1 integrates preferentially into active cellular genes in *in vitro* infections of transformed cell lines (35, 46). However, in an elegant cell line model of HIV-1 latency, integration sites were found in chromosomal regions disfavoring transcription, such as centromeric regions (27). This led to the notion that latency was determined primarily by the site of integration. However, the first analysis of HIV-1 integration sites in infected individuals found that HIV-1 proviruses integrated into transcriptionally active cellular genes in resting CD4⁺ T cells from patients on HAART (18). The same pattern has been observed in peripheral blood mononuclear cells (PBMC) from untreated individuals (36). In these studies, integrated proviruses were detected by specific PCR strategies, but the replication competence of the proviruses was not assessed. Given that only a small fraction of the integrated proviruses in resting CD4⁺ T cells from infected individuals appear to be capable of producing infectious virus following cellular activation (8), the possibility remained that integration sites had a different character in the subset of cells harboring replication-competent viral genomes. It has also been unclear whether the observed pattern reflects the likelihood of initial integration into particular regions or subsequent selection of cells carrying

* Corresponding author. Mailing address: Department of Medicine, Johns Hopkins University Bloomberg School of Public Health, Howard Hughes Medical Institute, 720 Rutland Avenue, Baltimore, MD 21205. Phone: (410) 955-2958. Fax: (443) 287-6218. E-mail: rsiliciano@jhmi.edu.

† Present address: National Taiwan University Hospital, Taipei, Taiwan.

[∇] Published ahead of print on 23 March 2011.

particular types of integration events. Nevertheless, the finding that HIV-1 genomes reside within cellular genes raised the possibility that an additional mechanism, transcriptional interference, could also play a role in latency.

Transcriptional interference is defined as direct suppression of one transcription unit by another in *cis* (48). Two adjacent promoters have been shown to transcriptionally interfere with each other through perturbation of the association of the transcription initiation complex, dislodgement of a sitting promoter-bound complex, or collision between transcription elongation complexes moving in opposite directions (1, 6, 38, 39, 44). Transcriptional interference has been demonstrated in an experimental system with two tandem HIV-1 long terminal repeats (LTRs) in HeLa cells (17). In addition, transcription of the 5' LTR has been shown to interfere with transcription initiating at the 3' LTR in Jurkat cells (34). In an elegant study in a cell line model, Lenasi et al. demonstrated that host gene transcription suppressed expression of a provirus integrated within the gene (34). Both positive and negative effects of host gene transcription on expression of integrated proviruses have been noted in studies on individual cell clones (20, 34). In a cell line model of HIV-1 latency, lower levels of HIV-1 gene expression have been observed when integration into highly expressed host genes occurs (35). Host gene transcription through the LTR upstream of the HIV-1 transcription start site has been detected both in resting CD4⁺ T cells from patients on HAART (18) and in primary CD4⁺ T cells infected *in vitro* (34). However, whether host gene transcription interferes with expression from an integrated provirus contained within that gene has not yet been demonstrated in any primary-cell system or *in vivo* study.

In an effort to more accurately model the *in vivo* regulation of HIV-1 gene expression, several groups have recently developed primary CD4⁺ T cell models of HIV-1 latency (4, 5, 37, 45, 62). The model developed by Yang et al. is of particular interest because it utilizes a prolonged culture period in cytokine-free media to allow infected T lymphoblasts to return to a fully quiescent state similar to that of resting memory CD4⁺ T cells *in vivo*. In the present study, we have used this primary-cell model to study the viral integration sites in latently infected resting CD4⁺ T cells that are capable of upregulating HIV-1 gene expression following cellular activation. Serial analysis of gene expression (SAGE) was carried out to determine transcription levels of host genes harboring integrated proviruses. By comparing viral integration sites and the levels of expression of the host gene, we have identified new features of the integration sites of latent HIV-1, which may provide a better understanding of the mechanism of HIV-1 latency.

MATERIALS AND METHODS

HIV-1 infection of Bcl-2-transduced primary CD4⁺ T cells. Peripheral blood for the isolation of primary CD4⁺ T cells was obtained from healthy adult volunteers. This study was approved by the Johns Hopkins Institutional Review Board. Written informed consent was provided by all study participants. Latently infected, Bcl-2-transduced primary CD4⁺ cells were generated as previously described (62). Briefly, CD4⁺ T cells were isolated from healthy donors and then costimulated by plate-bound anti-CD3 and soluble anti-CD28 antibodies, 100 U/ml interleukin-2 (IL-2), and T cell growth factor-enriched medium. Three days later, the cells were transduced with a Bcl-2-encoding lentiviral vector. The cells were allowed to return to a resting state, and then viable cells were isolated using Ficoll-Hypaque density gradient centrifugation.

Bcl-2-transduced cells were again costimulated as described above for 3 days and then infected with the HIV-1 reporter virus NL4-3-Δ6-drEGFP at a multiplicity of infection (MOI) of <0.1. NL4-3-Δ6-drEGFP has inactivating mutations in the *gag*, *vif*, *vpr*, *vpu*, and *nef* genes and a destabilized form of enhanced green fluorescent protein (EGFP) (Clontech) inserted into the *env* open reading frame (ORF). The virus was generated by cotransfecting 293T cells with NL4-3-Δ6-drEGFP, an X4 *env* expression vector (pCXCR4), and the helper plasmid pC-Help as described previously (40).

Analysis of HIV-1 integration sites. Genomic DNA from the three cell populations representing acute, persistent, and latent infection was isolated using a QIAamp DNA microkit (Qiagen). HIV-1 integration sites were determined as previously described (18). Briefly, genomic DNA was digested with PstI, and the products were diluted and ligated overnight to generate circularized DNA. Inverse PCR was carried out to amplify the junction between the HIV-1 5' LTR and upstream cellular DNA. Bands representing unique integration events were eluted and sequenced. Human genomic sequences were mapped using the UCSC Bioinformatics Human Genome database. An integration site was considered valid if a unique human genomic sequence was joined directly to the end of the 5' LTR of HIV-1.

Construction of SAGE library of Bcl-2-transduced CD4⁺ T cells. Uninfected, Bcl-2-transduced resting CD4⁺ cells were isolated by negative selection using mouse monoclonal antibodies CD25, CD69, and HLA-DR to deplete activated cells. Activated CD4⁺ T cells were collected 3 days after costimulation of uninfected Bcl-2-transduced CD4⁺ T cells with plate-bound anti-CD3 and soluble anti-CD28 antibodies. Total RNA was isolated from both activated and resting Bcl-2-transduced primary CD4⁺ T cells using an RNeasy minikit (Qiagen).

Seventeen-nucleotide (nt) tags from cellular transcripts were generated using a digital gene expression tag profiling preparation kit (Illumina, San Diego, CA). Briefly, mRNA was purified from 1 μg of total RNA using Oligo-dT magnetic beads (Invitrogen). After first- and second-strand cDNA synthesis, the double-stranded cDNA was digested with NlaIII and ligated to an adapter containing an MmeI restriction site. After MmeI digestion, the 17-nt tag of cellular transcript joined with adapter was released from Oligo-dT magnetic beads and ligated to a second adapter. The ligated product was then amplified by 18 cycles of PCR. The 85-bp DNA fragment was purified from a polyacrylamide gel and sequenced on a genome analyzer system (Illumina, San Diego, CA).

SAGE library analysis. All 17-nt tags were annotated using the "long best gene" database provided by <http://cgap.nci.nih.gov/SAGE/>. Potentially erroneous sequences were removed from the tag library if one of the following conditions applied: (i) genes were differentially expressed between two replicates, (ii) genes appeared only once, or (iii) genes were 1 bp different from more highly expressed genes unless mapped to the genome or transcriptome database (41). After data filtration, frequencies of identical tags in two replicates were combined. The combined frequency of unique tags was converted into relative abundance as follows: relative abundance = tag frequency × 10⁶/total tag number.

RESULTS

Generation of acutely infected, persistently infected, and latently infected CD4⁺ T cell populations. To study the role of host gene transcription in HIV-1 latency, primary CD4⁺ T cells from healthy donors were transduced with a Bcl-2-encoding lentiviral vector to allow their long-term *in vitro* survival and then infected with a previously described, X4-pseudotyped HIV-1 reporter virus encoding GFP (NL4-3-Δ6-drEGFP). The strategy for generating acutely infected, persistently infected, and latently infected populations of primary CD4⁺ T cells is shown in Fig. 1A. Acutely infected cells were collected 3 days postinfection regardless of GFP expression. For the persistently infected and latently infected populations, cells were further cultured in medium without exogenous cytokines for 3 to 4 weeks. Cell sorting was carried out to separate GFP-positive and GFP-negative populations. The GFP-positive population was considered persistently infected. The GFP-negative population was considered to contain latently infected cells. Importantly, there was no spontaneous GFP reexpression when GFP-negative cells were maintained in medium without exogenous cytokines (data not shown). Then, the GFP-nega-

FIG. 1. Generation of latently infected primary CD4⁺ T cells. (A) Experimental approach to generating acutely infected, persistently infected, and latently infected CD4⁺ T cells. Activated Bcl-2-transduced CD4⁺ T cells were infected with NL4-3-Δ6-drEGFP. Sorting for GFP-positive cells was carried out 25 days postinfection. The sorted GFP-negative cells were cultured for 3 days and then further costimulated by anti-CD3 and anti-CD28 antibodies. Cells derived from the GFP-negative population that upregulated GFP following cellular activation were used to define integration sites in latently infected cells. Roman numerals I to IV refer to the populations analyzed by flow cytometry as shown in panel B. (B) Flow cytometric analysis of GFP expression in cell populations at various stages (I to IV) in the protocol for generating latently infected cells. (C) Bcl-2-transduced CD4⁺ T cells were infected with NL4-3-Δ6-drEGFP. Raltegravir (10 μM) was added 3 days after infection and was maintained throughout the whole experiment. GFP-negative cells were sorted 7, 14, and 21 days after infection and were costimulated with anti-CD3 and anti-CD28 antibodies. GFP expression was measured by fluorescence-activated cell sorter analysis 2 days after costimulation. FL2-H, red fluorescence.

FIG. 2. Genomic distribution and orientation of HIV-1 integration sites. (A) Nature of integration sites in acutely infected, persistently infected, and latently infected resting CD4⁺ T cells. HIV-1 integration sites were determined by inverse PCR. Intron and exon boundaries were analyzed using Refseq, EST, Human mRNA, and Unigene databases. ITG, intergenic space; other, integration sites matched with other gene or transcription databases but not included in Refseq. (B) Comparison of the transcriptional orientations of integrated proviruses with respect to host genes in acutely, persistently, and latently infected cells. (C) Comparison of HIV-1 orientations with respect to host genes in latently infected cells from three independent infections. (D) Comparison of HIV-1 transcriptional orientations for integration sites in repeatedly targeted genes in persistently and latently infected populations. *P* values shown in panels B and C were calculated using 1:1 as the expected ratio. The *P* value shown in panel D was calculated by direct comparison between persistently and latently infected populations.

tive cells were cultured for another 3 days and then restimulated by plate-bound anti-CD3 and soluble anti-CD28 antibodies. Sorting was then carried out to isolate those cells that were induced to express GFP after stimulation. These inducible cells were considered to represent latently infected CD4⁺ T cells. Three independent NL4-3-Δ6-drEGFP infections of Bcl-2-transduced primary CD4⁺ T cells from the same donor were carried out. Flow cytometric analysis of GFP expression at different stages of a representative infection is shown in Fig. 1B.

It is unclear when viruses enter into a latent state in our model. To address this question, GFP[−] cells were isolated at days 7, 14, and 21 after infection and then stimulated with anti-CD3 and anti-CD28 antibodies. As shown in Fig. 1C, latent viruses were detected as early as day 7. The frequency of latently infected cells did not increase during the following 2 weeks. These results indicate that latency is established rapidly

in this system. Infected cells that do not enter a latent state during the first week continue to express viral genes, possibly due to an effective Tat-dependent positive-feedback loop (57).

HIV-1 integration sites in primary CD4⁺ T cells. Genomic DNA was isolated from the acutely infected, persistently infected, and latently infected populations, and integration sites were determined by inverse PCR. Sites were mapped using the UCSC Bioinformatics Human Genome database. As shown in Fig. 2A, HIV-1 was found preferentially integrated into transcription units in acutely infected cells, consistent with previous results (46). In acutely infected cells, 91% of the integration events were within transcriptional units. In persistently and latently infected cell populations, 87% and 94% of integrated proviruses, respectively, were found in transcriptional units. Integration into intergenic regions was uncommon (6%) in latently infected cells (Fig. 2A). Taken together, these results suggest that the integration patterns observed in resting CD4⁺

T cells from patients (18) and in latently infected cells in this primary-cell model are very similar. Previous *in vivo* observations did not determine whether resting CD4⁺ T cells harboring integrated viral genomes could actually upregulate virus gene expression following cellular activation. Our current findings thus extend previous work by showing that in latently infected primary CD4⁺ T cells that can upregulate HIV-1 gene expression following cellular activation, the viral genomes are integrated into genes that are actively transcribed in resting CD4⁺ T cells (Fig. 3B). Thus, this study provides further evidence against the notion that latency results from initial integration into chromosomal regions that are intrinsically repressive for transcription. In addition, because viral genomes preferentially reside in active cellular genes in both acutely and latently infected cells, it is possible that the bias toward active genes rather than intergenic regions is established during the initial integration in activated cells and is not altered by the subsequent events that lead to the establishment of latency in some infected cells.

Orientation preference of latent proviruses. We also examined the transcriptional orientation of integrated proviruses with respect to the host gene. No integration preference was detected in a previous study of integration events in resting CD4⁺ T cells from patients on HAART (18). However, in that study, it was not possible to ascertain whether the integrated proviruses were inducible. In the system described here, we examined integration sites in latently infected cells that can upregulate HIV-1 gene expression following stimulation. In acutely infected and persistently infected cells, differences between the numbers of integration sites in the same versus convergent orientation with respect to the host gene were not statistically significant. However, HIV-1 was found preferentially in the same transcriptional orientation as the host gene in latently infected cells (Fig. 2B). This bias for the same versus convergent orientation (63.8% versus 36.2%) was statistically significant ($P < 0.0001$). These results represent aggregated data from three independent preparations of latently infected cells. The same statistically significant orientation preference was observed when each of these independent preparations of latently infected cells was analyzed separately (Fig. 2C).

We compared the integration sites in persistently and latently infected cells and found 29 host genes which had at least one integration site detected in both populations. In each case, the integration sites were at different positions within the same gene. The detection of distinct integration events in the same genes in persistently infected and latently infected cell populations allowed us to confirm the orientation bias described above with an internally controlled data set. As shown in Fig. 2D, within the same set of targeted genes, latent proviruses were found integrated preferentially in the same orientation while persistently expressed proviruses preferred the convergent orientation ($P = 0.008$).

Taken together, these results demonstrate that there is a subtle but significant orientation preference in latently infected cells, indicating that transcriptional orientation in the same direction as the host gene is a favored condition contributing to proviral latency. These results provide direct evidence that host gene transcription affects latency in primary resting CD4⁺ T cells.

HIV-1 integration favored in highly active genes. Previous studies in cell line systems have described a relationship between HIV-1 latency and the level of expression of host gene as measured by microarray analysis (35). While microarray analysis allows comparison between the levels of expression of a given gene under different conditions, it provides only a crude measure of relative expression level of different genes. To further study the influence of host gene transcription on HIV-1 latency, serial analysis of gene expression (SAGE) was carried out to assess the level of expression of host genes in which integration sites were found in primary CD4⁺ T cells. In comparison to microarray analysis, SAGE has the advantage of providing a quantitative measure of differences in the levels of expression of different genes in the same cell (54, 56). Replicate samples of uninfected, Bcl-2-transduced resting and activated CD4⁺ T cells were prepared from the same healthy donor used for integration site analysis. RNA was isolated and processed for SAGE. After data filtration to remove erroneous sequences and matching with the "long best gene" database, a total of 20,643 and 22,279 unique tags were acquired from activated and resting cells, respectively. The frequency of a unique tag in one million tags, termed relative abundance, was used to represent the transcription level of the corresponding gene. For most transcriptionally active genes, the frequencies of SAGE tags detected were similar in resting and activated cells. These genes fall on the diagonal on a plot of log transcription levels in resting versus activated T cells (Fig. 3A). Since the total amount of RNA in activated Bcl-2-transduced CD4⁺ T cells was 5- to 10-fold greater than that in resting Bcl-2-transduced CD4⁺ T cells, these results indicate that gene transcription is globally and proportionally downregulated when cells transition from an activated to a resting state. Genes known to be upregulated in resting CD4⁺ T cells (e.g., IL-7 receptor gene) or activated CD4⁺ T cells (e.g., thymidylate synthetase and chemokine ligand 2 genes) gave the expected patterns, while housekeeping genes (e.g., ribosomal protein and β -actin genes) showed no differential up- or downregulation.

The transcription levels of host genes harboring integrated proviruses in acutely, persistently, or latently infected cells, as well as in 192 randomly selected genes, were determined from the SAGE data. Integration sites in intergenic regions were excluded from the analysis. Integration sites in genes for which no SAGE tags were found were considered to be in transcriptionally inactive genes. In persistently and latently infected resting CD4⁺ T cells, fewer than 8% of genes containing integration sites were transcriptionally inactive whereas 54% of randomly selected genes were transcriptionally inactive (Fig. 3B). In addition, those genes into which HIV-1 had integrated had levels of transcription that were 10- to 100-fold higher than levels of transcription in randomly selected genes in resting CD4⁺ T cells (Fig. 3C). Only 6% of genes have transcription levels higher than 2 log in this analysis, and therefore integration into these rare, highly expressed genes is less common. Taken together, these results demonstrate that latent viral genomes are frequently found in genes which have moderate to high levels of expression in resting CD4⁺ T cells.

HIV-1 replicates preferentially in activated CD4⁺ T cells, and viral latency can be established when activated cells revert to the resting memory state and shut down viral gene tran-

FIG. 3. Latent HIV-1 proviruses are found in highly active genes in resting and activated Bcl-2-transduced CD4⁺ T cells. (A) Levels of transcription of host genes in resting and activated cells. The relative abundance of transcripts of different host genes was measured by SAGE. Data are displayed on a logarithmic scale. Several known genes are shown as examples. (B) Percentages of integration sites in transcriptionally inactive genes in acutely infected, persistently infected, and latently infected cell populations. For purposes of comparison, the percentage of transcriptionally inactive genes among 192 genes randomly selected from Unigene database is shown. Levels of gene expression were determined by SAGE using RNA from uninfected, resting Bcl-2-transduced CD4⁺ T cells. (C) Levels of transcription in genes containing integration sites. The log scale of gene transcription levels was divided into nine bins as follows: 0 to 0.25, 0.25 to 0.75, 0.75 to 1.25, 1.25 to 1.75, 1.75 to 2.25, 2.25 to 2.75, 2.75 to 3.25, 3.25 to 3.75, and 3.75 to 4.25. The percentage of genes within each bin is shown. (D) Transcription levels in uninfected, resting, and activated Bcl-2-transduced cells of randomly selected genes or genes in which integration sites were found in the acutely infected, persistently infected, and latently infected populations. Side bars represent the numbers of integration sites within the corresponding regions.

FIG. 4. HIV-1 orientation preference at different host gene transcription levels. The gene transcription scale was equally divided into eight bins. Among the total integration sites from the indicated populations, each bar represents the percentage of sites with the indicated orientation located in genes with the indicated level of expression in resting or activated cells. *P* values were calculated using 1:1 as the expected ratio. Only significant *P* values are shown (*, *P* < 0.05; **, *P* < 0.01).

scription (19, 62). Therefore, it was of interest to investigate the host gene transcription level in activated cells because the virus initially integrates into the genome of activated cells and because the level of transcription of viral genes is influenced by the local chromosomal environment (28). We measured the levels of host gene transcripts in activated, uninfected Bcl-2-transduced CD4⁺ T cells using SAGE and compared the results with the data from resting, uninfected Bcl-2-transduced cells. In both resting and activated CD4⁺ T cells, most of the randomly selected transcriptionally active genes are transcribed at relatively low levels (Fig. 3D). In contrast, those genes in which HIV-1 integration sites were detected in the acutely infected, persistently infected, and latently infected populations had high levels of transcription in both resting and activated states (Fig. 3D). The above-described analysis of the transcription level of host genes in which HIV-1 integration had occurred was based on SAGE of uninfected Bcl-2-transduced CD4⁺ T cells. To confirm that latent HIV-1 infection did not change the transcription level of these several hundred host genes in resting CD4⁺ T cells, we compared the levels of transcription of these genes in latently infected and uninfected populations of Bcl-2-transduced resting CD4⁺ T cells by microarray analysis and found no significant difference (data not shown). The results demonstrated that the SAGE data from uninfected CD4⁺ T cells properly represented the gene transcription in latently infected resting CD4⁺ T cells.

Taken together, these results suggest that the latency in CD4⁺ T cells does not result from specific integration into a chromosomal environment that is repressive for transcription. The fact that viral genomes are found in actively transcribed host genes in acute, persistent, and latent infection is consistent with the idea that integration occurs in activated T cells at genes that are transcriptionally active and therefore accessible. As the cells revert to a resting state, the transcriptional environment becomes less favorable for HIV-1 gene expression

even though most of the integration sites are located in genes that are actively expressed in resting CD4⁺ T cells.

Transcription levels and the orientation preference of latent HIV-1. As discussed above, latent HIV-1 proviruses were found preferentially in the same transcriptional orientation as the host gene. We further investigated whether the level of transcription correlated with orientation preference. No orientation preference was observed in acutely infected or persistently infected populations at any transcription level in either activated or resting CD4⁺ T cells (Fig. 4). In the latently infected population, the ratio between same orientation and convergent orientation was greater in genes with moderate- to high-level expression (Fig. 4). This preference was not seen in genes that had low-level expression, probably because the inefficient host gene transcription had no significant influence on HIV-1 transcription. Together, these results suggest that orientation preference of latent HIV-1 is correlated with host gene transcription level, providing further evidence that host gene expression affects latency.

DISCUSSION

A previous study using resting CD4⁺ T cells from patients on HAART showed that HIV-1 genomes were preferentially integrated into actively transcribed host genes (18). However, the majority of proviruses in resting CD4⁺ T cells cannot be induced to release replication-competent viruses following cellular activation (8). Thus, it is possible that the nature of integration sites of replication-competent latent HIV-1 is different from that observed for the whole population of integrated proviruses. In this study, we analyzed HIV-1 integration sites in primary CD4⁺ T cells capable of upregulating HIV-1 gene expression following cellular activation. We showed that more than 90% of the latent proviruses were integrated within actively transcribed host genes, most of which were transcribed

FIG. 5. Transcriptional interference from upstream host gene transcription in the same and convergent orientations. (Top) The provirus lies in the same orientation with respect to host gene transcription. Transcription complexes initiating at the host gene promoter consistently read through the viral promoter region and inhibit the initiation or elongation of viral transcription. Polymerase complexes transcribing the host gene may accumulate at the 5' LTR due to the presence of a polyadenylation site and block access of initiation complexes to the LTR. (Bottom) The provirus lies in the convergent orientation with respect to host gene transcription. Both virus and host gene transcriptionally interfere with each other. Transcription from the viral 3' LTR might prevent host gene read-through and protect transcriptional activity from the 5' LTR.

at high levels in uninfected cells. These results suggest that in primary resting $CD4^+$ T cells, HIV-1 can stably maintain a state of latency even within highly active genes and can be reactivated upon stimulation. Integration into the repressive chromosomal region is not necessary for transcriptional silencing of integrated proviruses.

Integrated HIV-1 was found equally in same and convergent orientations relative to the direction of host gene transcription in both acutely infected (46) and persistently infected (18, 35, 53) $CD4^+$ T cells. Surprisingly, we found that the same orientation was significantly preferred in latently infected cells. Since this orientation preference was not observed in acutely or persistently infected cells in our experiments, the possibility that the orientation preference was the result of the integration site detection assays can be excluded. The fact that an orientation preference was not seen in cell line models indicates that viral latency established in cell line models is at least partially different from that in primary cells. Our results suggest that same orientation provides a more repressive local chromosomal environment for viral transcription than convergent orientation. However, approximately one-third of latent proviruses were in the convergent orientation, indicating that this orientation preference is caused by a modest effect(s) which contributes only partially to the establishment of latency.

It is also important to point out that an orientation bias has not yet been observed in cells harboring replication-competent

virus in patients. This reflects the technical difficulty in simultaneously assessing replication competence and integration site orientation in individual cells from patients. A further caveat with our *in vitro* primary-cell model is the possibility that overexpression of Bcl-2 may have some unanticipated effect on the mechanisms involved in latency. In addition, the model makes use of a virus with inactivating mutations in all viral ORFs except *tat* and *rev*. This was done to decrease viral cytopathic effects and increase the yield of latently infected cells. It is possible that viral gene products could modify the state of the host cell in a way that affects the establishment of latency.

Among all currently known mechanisms which maintain viral latency, only RNA interference and transcriptional interference can possibly explain orientation preferences. In the convergent orientation, viral transcripts and host gene read-through products could in principle form a double-strand RNA duplex, which most likely would have a negative influence on viral gene expression and could facilitate the establishment of viral latency in convergent orientation. This hypothetical effect from RNA interference favors the establishment of latency for proviruses with convergent orientation. Since the opposite preference was observed in our primary-cell model, we conclude that double-strand RNA, if produced, has a minor effect on viral gene expression. We cannot exclude the possibility that

double-strand RNA may turn on cellular pathways and indirectly influence viral gene transcription.

Transcriptional interference between adjacent promoters is also reported to control viral transcription (17, 20, 34). When proviruses are integrated in the same orientation as the host gene, transcription complexes initiating at the promoter of the upstream host gene can read through the viral promoter and inhibit viral transcription by dislodging prebound transcription initiation/elongation complexes assembled on the LTR or preventing the transcription initiation complex from binding to the LTR. The poly(A) signal sequence in the R region of the 5' LTR can cause termination of host gene transcription and accumulation of transcription elongation complexes on the LTR, which further inhibits viral transcription initiation, but cannot prevent host gene transcription complexes from reading through the viral promoter region. Therefore, when virus integrates in the same orientation as host gene transcription, the viral promoter is always under negative control by the host promoter, while host promoter activity is not directly influenced by the virus. In the convergent orientation, the host promoter and the viral promoter interfere with each other. Moreover, the U3 region in the 3' LTR can also initiate transcription, which can partially prevent the host gene read-through and protect transcription from the 5' LTR (Fig. 5). These considerations may explain why latently infected cells more commonly show this orientation preference.

Overall, our experiments suggest that host gene transcription level and orientation are involved in the establishment of HIV-1 latency. Viral transcription is more efficiently inhibited when virus lies in the same orientation as the direction of host gene transcription. Establishment of HIV-1 latency is believed to be a multifactorial process. The state of latency is maintained due to the balance among many factors. Disrupting any single factor can break the balance and result in the reactivation of latent HIV-1. Inhibition of transcription interference from host genes provides a potential mechanism to disrupt latency.

ACKNOWLEDGMENTS

This work was supported by NIH grant AI43222 and by the Howard Hughes Medical Institute.

REFERENCES

- Adhya, S., and M. Gottesman. 1982. Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**:939–944.
- Blazkova, J., et al. 2009. CpG methylation controls reactivation of HIV from latency. *PLoS Pathog.* **5**:e1000554. doi:10.1371/journal.ppat.1000554.
- Bohnlein, E., et al. 1988. The same inducible nuclear proteins regulates mitogen activation of both the interleukin-2 receptor- α gene and type 1 HIV. *Cell* **53**:827–836.
- Bosque, A., and V. Planelles. 2009. Induction of HIV-1 latency and reactivation in primary memory CD4⁺ T cells. *Blood* **113**:58–65. doi:10.1182/blood-2008-07-168393.
- Brooks, D. G., S. G. Kitchen, C. M. Kitchen, D. D. Scripture-Adams, and J. A. Zack. 2001. Generation of HIV latency during thymopoiesis. *Nat. Med.* **7**:459–464. doi:10.1038/86531.
- Callen, B. P., K. E. Shearwin, and J. B. Egan. 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* **14**:647–656. doi:10.1016/j.molcel.2004.05.010.
- Chun, T. W., et al. 1995. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nat. Med.* **1**:1284–1290.
- Chun, T. W., et al. 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**:183–188. doi:10.1038/387183a0.
- Chun, T. W., et al. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* **94**:13193–13197.
- Chun, T. W., et al. 2003. Gene expression and viral production in latently infected, resting CD4⁺ T cells in viremic versus aviremic HIV-infected individuals. *Proc. Natl. Acad. Sci. U. S. A.* **100**:1908–1913. doi:10.1073/pnas.0437640100.
- Coull, J. J., et al. 2000. The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J. Virol.* **74**:6790–6799.
- Cujec, T. P., et al. 1997. The HIV transactivator TAT binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**:2645–2657.
- Duh, E. J., W. J. Maury, T. M. Folks, A. S. Fauci, and A. B. Rabson. 1989. Tumor necrosis factor α activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF- κ B sites in the long terminal repeat. *Proc. Natl. Acad. Sci. U. S. A.* **86**:5974–5978.
- Finzi, D., et al. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**:1295–1300.
- Finzi, D., et al. 1999. Latent infection of CD4⁺ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**:512–517. doi:10.1038/8394.
- Ganesh, L., et al. 2003. The gene product Murr1 restricts HIV-1 replication in resting CD4⁺ lymphocytes. *Nature* **426**:853–857. doi:10.1038/nature02171.
- Greger, I. H., F. Demarchi, M. Giacca, and N. J. Proudfoot. 1998. Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Res.* **26**:1294–1301.
- Han, Y., et al. 2004. Resting CD4⁺ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J. Virol.* **78**:6122–6133. doi:10.1128/JVI.78.12.6122-6133.2004.
- Han, Y., M. Wind-Rotolo, H. C. Yang, J. D. Siliciano, and R. F. Siliciano. 2007. Experimental approaches to the study of HIV-1 latency. *Nat. Rev. Microbiol.* **5**:95–106. doi:10.1038/nrmicro1580.
- Han, Y., et al. 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe* **4**:134–146. doi:10.1016/j.chom.2008.06.008.
- He, G., and D. M. Margolis. 2002. Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. *Mol. Cell. Biol.* **22**:2965–2973.
- Hermankova, M., et al. 2003. Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4⁺ T lymphocytes in vivo. *J. Virol.* **77**:7383–7392.
- Herrmann, C. H., and A. P. Rice. 1995. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J. Virol.* **69**:1612–1620.
- Huang, J., et al. 2007. Cellular microRNAs contribute to HIV-1 latency in resting primary CD4⁺ T lymphocytes. *Nat. Med.* **13**:1241–1247. doi:10.1038/nm1639.
- Jones, K. A., and B. M. Peterlin. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**:717–743. doi:10.1146/annurev.bi.63.070194.003441.
- Joos, B., et al. 2008. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc. Natl. Acad. Sci. U. S. A.* **105**:16725–16730. doi:10.1073/pnas.0804192105.
- Jordan, A., D. Bisgrove, and E. Verdin. 2003. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**:1868–1877. doi:10.1093/emboj/cdg188.
- Jordan, A., P. Defechereux, and E. Verdin. 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* **20**:1726–1738. doi:10.1093/emboj/20.7.1726.
- Kao, S. Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin. 1987. Antitermination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**:489–493. doi:10.1038/330489a0.
- Kauder, S. E., A. Bosque, A. Lindqvist, V. Planelles, and E. Verdin. 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog.* **5**:e1000495. doi:10.1371/journal.ppat.1000495.
- Kinoshita, S., et al. 1997. The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. *Immunity* **6**:235–244.
- Lassen, K., Y. Han, Y. Zhou, J. Siliciano, and R. F. Siliciano. 2004. The multifactorial nature of HIV-1 latency. *Trends Mol. Med.* **10**:525–531. doi:10.1016/j.molmed.2004.09.006.
- Lassen, K. G., K. X. Ramyar, J. R. Bailey, Y. Zhou, and R. F. Siliciano. 2006. Nuclear retention of multiply spliced HIV-1 RNA in resting CD4⁺ T cells. *PLoS Pathog.* **2**:e68. doi:10.1371/journal.ppat.0020068.
- Lenasi, T., X. Contreras, and B. M. Peterlin. 2008. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe* **4**:123–133. doi:10.1016/j.chom.2008.05.016.
- Lewinski, M. K., et al. 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**:6610–6619. doi:10.1128/JVI.79.11.6610-6619.2005.

36. Liu, H., et al. 2006. Integration of human immunodeficiency virus type 1 in untreated infection occurs preferentially within genes. *J. Virol.* **80**:7765–7768. doi:10.1128/JVI.00542-06.
37. Marini, A., J. M. Harper, and F. Romerio. 2008. An in vitro system to model the establishment and reactivation of HIV-1 latency. *J. Immunol.* **181**:7713–7720.
38. Martens, J. A., L. Laprade, and F. Winston. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**:571–574. doi:10.1038/nature02538.
39. Mazo, A., J. W. Hodgson, S. Petruk, Y. Sedkov, and H. W. Brock. 2007. Transcriptional interference: an unexpected layer of complexity in gene regulation. *J. Cell Sci.* **120**:2755–2761. doi:10.1242/jcs.007633.
40. Mochizuki, H., J. P. Schwartz, K. Tanaka, R. O. Brady, and J. Reiser. 1998. High-titer human immunodeficiency virus type 1-based vector systems for gene delivery into nondividing cells. *J. Virol.* **72**:8873–8883.
41. Morrissy, A. S., et al. 2009. Next-generation tag sequencing for cancer gene expression profiling. *Genome Res.* **19**:1825–1835. doi:10.1101/gr.094482.109.
42. Nabel, G., and D. Baltimore. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**:711–713. doi:10.1038/326711a0.
43. Pomerantz, R. J., D. Trono, M. B. Feinberg, and D. Baltimore. 1990. Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. *Cell* **61**:1271–1276.
44. Prescott, E. M., and N. J. Proudfoot. 2002. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* **99**:8796–8801. doi:10.1073/pnas.132270899.
45. Sahu, G. K., et al. 2006. A novel in vitro system to generate and study latently HIV-infected long-lived normal CD4⁺ T-lymphocytes. *Virology* **355**:127–137. doi:10.1016/j.virol.2006.07.020.
46. Schroder, A. R., et al. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**:521–529.
47. Selby, M. J., and B. M. Peterlin. 1990. Trans-activation by HIV-1 Tat via a heterologous RNA binding protein. *Cell* **62**:769–776.
48. Shearwin, K. E., B. P. Callen, and J. B. Egan. 2005. Transcriptional interference—a crash course. *Trends Genet.* **21**:339–345. doi:10.1016/j.tig.2005.04.009.
49. Siliciano, J. D., et al. 2003. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4⁺ T cells. *Nat. Med.* **9**:727–728. doi:10.1038/nm880.
50. Strain, M. C., et al. 2003. Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence. *Proc. Natl. Acad. Sci. U. S. A.* **100**:4819–4824. doi:10.1073/pnas.0736332100.
51. Tyagi, M., R. J. Pearson, and J. Karn. 2010. Establishment of HIV latency in primary CD4⁺ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. *J. Virol.* **84**:6425–6437. doi:10.1128/JVI.01519-09.
52. Van Lint, C., S. Emiliani, M. Ott, and E. Verdin. 1996. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J.* **15**:1112–1120.
53. Vatakis, D. N., S. Kim, N. Kim, S. A. Chow, and J. A. Zack. 2009. Human immunodeficiency virus integration efficiency and site selection in quiescent CD4⁺ T cells. *J. Virol.* **83**:6222–6233. doi:10.1128/JVI.00356-09.
54. Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. *Science* **270**:484–487.
55. Verdin, E., P. Paras, Jr., and C. Van Lint. 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J.* **12**:3249–3259.
56. Wang, S. M. 2007. Understanding SAGE data. *Trends Genet.* **23**:42–50. doi:10.1016/j.tig.2006.11.001.
57. Weinberger, L. S., J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer. 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**:169–182. doi:10.1016/j.cell.2005.06.006.
58. West, M. J., A. D. Lowe, and J. Karn. 2001. Activation of human immunodeficiency virus transcription in T cells revisited: NF-kappaB p65 stimulates transcriptional elongation. *J. Virol.* **75**:8524–8537.
59. Williams, S. A., et al. 2006. NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.* **25**:139–149. doi:10.1038/sj.emboj.7600900.
60. Williams, S. A., and W. C. Greene. 2007. Regulation of HIV-1 latency by T-cell activation. *Cytokine* **39**:63–74. doi:10.1016/j.cyt.2007.05.017.
61. Wong, J. K., et al. 1997. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**:1291–1295.
62. Yang, H. C., et al. 2009. Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. *J. Clin. Invest.* **119**:3473–3486. doi:10.1172/JCI39199.
63. Zhang, L., et al. 2000. Genetic characterization of rebounding HIV-1 after cessation of highly active antiretroviral therapy. *J. Clin. Invest.* **106**:839–845. doi:10.1172/JCI10565.