Genotypic and Phenotypic Characterization of Long Terminal Repeat Sequences from Long-Term Survivors of Human Immunodeficiency Virus Type 1 Infection

LINQI ZHANG,1 YAOXING HUANG,1 HANNAH YUAN,1 BENJAMIN K. CHEN,2 JAMES IP,1 AND DAVID D. HO1*

The Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10016,1 and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 021392

Received 4 December 1996/Accepted 2 April 1997

Human immunodeficiency virus type 1 (HIV-1)-infected individuals who remain asymptomatic despite prolonged infection present a unique opportunity to understand virologic and immunologic factors involved in the pathogenesis of AIDS. We have previously identified a group of long-term survivors (LTS) who are clinically healthy and immunologically normal despite 15 to 15 years of HIV-1 infection. In this study, we examined the 5′ long terminal repeat (5′ LTR) sequences in eight of these LTS. A total of 29 nucleotide sequences were obtained from their peripheral blood mononuclear cells (PBMC). Analysis of these sequences revealed no gross deletions within the 5′ LTR. Seven of the eight subjects shared nearly identical consensus sequences in the binding sites for NF-kB, Sp1, and the viral trans-activator Tat. In multiple samples from one individual (Pt 5), however, G-to-A hypermutations were found throughout the entire region, suggesting a genetically defective 5′ LTR. The effects of the observed genetic variations on LTR transcription were studied by transient transfection of an LTR-driven luciferase reporter gene and by infection with a full-length recombinant HIV-1 containing a luciferase reporter (HIVHXBLTRluc). A wide range of basal and Tat-induced transcriptional activities was found among the 5′ LTR from seven of the eight LTS in both transfected 293 cells and donor PBMC, suggesting a functionally intact 5′ LTR in these individuals. It is therefore unlikely that defects in the 5′ LTR are the underlying explanation for the benign clinical course associated with these seven individuals. However, functional abnormalities were found in the LTR from Pt 5 in directing both heterologous and viral gene expression, providing a possible genetic explanation for the low viral load and prolonged asymptomatic state of this individual. Last, a similar overall degree of genetic diversity was found among viruses from the LTS compared to those from patients with AIDS, reinforcing the notion that a strong correlation between the degree of genetic diversity and the rate of disease progression is unlikely.

Human immunodeficiency virus type 1 (HIV-1) infection leads to generalized dysfunction of the immune system and development of AIDS in the majority of HIV-1-infected individuals. There is a wide spectrum of progression rates from the initial infection to full-blown AIDS among infected individuals; some develop AIDS within months of infection (18, 31), while others remain clinically healthy for more than 10 years (1, 3, 4, 25, 28, 29, 37, 42). Currently, it is not clear whether all HIV-infected individuals will ultimately develop AIDS. The incubation time from the onset of infection to the development of complications is generally long, as in many lentivirus infections (1, 3, 4, 25, 28, 29, 37, 42). It has become evident recently from several cohort studies that a small proportion (approximately 5 to 8%) of infected individuals has remained clinically healthy and without evidence of immunodeficiency despite prolonged HIV-1 infection (1, 3, 4, 6, 25, 30, 37, 42). These individuals are, therefore, referred to as long-term survivors (LTS) or long-term nonprogressors. The explanation for the benign course of HIV-1 infection in these LTS may be related to viral or host factors or both (1, 3, 4, 6, 16, 17, 19, 23, 25, 28, 29, 33, 37, 42). Considerable research efforts have been made to elucidate the possible factors that may contribute to this attenuated clinical course (1, 3, 4, 6, 16, 17, 19, 23, 25, 28, 29, 33, 37, 42).

We and others have previously identified several groups of LTS (1, 3, 4, 6, 16, 17, 19, 23, 25, 28, 29, 33, 37, 42). Virologic and immunologic characterization of these individuals revealed several key findings. Levels of HIV-1 RNA in plasma and the viral burden in peripheral blood mononuclear cells (PBMC) were significantly lower in LTS than in progressors (1, 3, 4, 6, 16, 17, 19, 23, 25, 28, 29, 33, 37, 42). In many cases, infectious HIV-1 could not be isolated from PBMC and plasma samples (4, 6, 19, 33, 37), raising the possibility of an attenuated virus. In addition, broadly reactive neutralizing antibodies and potent cytotoxic T-cell activity against HIV-1 were detected in the majority of the study subjects (4, 6, 19, 23, 40). Also, the architecture of the lymph nodes of LTS appeared to remain relatively intact (37).

Efforts have also been made to characterize the genetic features of the viruses present in LTS. Sequence analyses of viral genomes in LTS have revealed genetic defects in multiple regions, including vif, vpr, vpu, tat, rev, and nef (6, 19, 23, 33). Perhaps the strongest report comes from the recent study of an Australian cohort of LTS infected with HIV-1 through blood transfusion from a single donor (6, 25). Multiple deletions in nef as well as in the U3 region of the 3′ long terminal repeat (LTR) are found in all members of this cohort (6, 25). The lack of disease in these individuals raises the possibility that the specific genetic defect in HIV-1 accounts for the prolonged asymptomatic state. This finding is remarkably similar to those in macaques that remain asymptomatic for a prolonged period of time after infection with the nef-deleted simian immuno-
ficiency virus strain SIVmac239 (21). In our cohort of LTS, however, nef was found to be both genetically and functionally intact (16, 17). In addition, detailed sequence analyses of vif, vpr, and vpu from our cohort of LTS revealed the integrity of these genes (44), indicating that a single common genetic determinant responsible for the well-being of LTS is unlikely. The functional properties of the env genes from our cohort of LTS were also studied. Envelope glycoproteins from four members of the cohort have been shown to be poorly processed and failed to mediate infection in pseudotyped HIV-1 virions, suggesting functional abnormalities in these Env glycoproteins (5).

In the present study, we have examined the genotypic and phenotypic properties of the HIV-1 LTR of the same cohort of LTS to further explore the possibility that specific viral defects account for the long-term survival of these individuals. Like all retroviruses, the HIV-1 provirus contains two identical LTRs, flanking the three structural genes, gag, pol, and env, that are essential for viral replication. Each LTR consists of three regions, namely U3, R, and U5 (28). HIV-1 transcription is regulated by multiple viral and cellular transcription factors that bind to distinct sequence elements in the 5′ LTR (8–10, 12, 15, 20, 39). For example, efficient transcription of HIV-1 in T lymphocytes or in macrophages requires the presence of the TATA box, the binding sites for NF-κB and Sp1, and the Tat-response region (TAR) (8–10, 12, 20, 27, 35, 36, 38, 41). Mutagenesis of these sequence elements either individually or collectively inhibits or activates HIV-1 replication (8–10, 12, 20, 27, 35, 36, 38, 41). It is generally accepted that the initiation of HIV-1 transcription is under the control of cellular factors, whereas the subsequent activation of HIV-1 transcription is largely mediated by virus-encoded trans-activators such as the Tat protein (8–10, 28, 36, 39).

To study the genotypic and phenotypic features of the HIV-1 LTR, PCR amplification, DNA sequencing and sequence analyses, and in vitro functional analyses of the 5′ LTRs were performed on the samples collected from eight of our LTS. Our results suggest the following conclusions. First, no gross sequence deletions within 5′ LTRs were performed on the samples collected from eight of our LTS. But within each subject, there are no gross sequence deletions or insertions in any of the study subjects. Within each sample, the PCR products appeared homogeneous in length as reflected by a single, sharp band on the polyacrylamide gel (data not shown). In multiple, independent PCR amplifications these findings remained consistent. Initially, efforts had also been made to use reverse transcription-PCR to amplify viral RNA sequences from the collected plasma samples. But due to the low viral loads in these individuals, the attempts failed despite repeated efforts. However, 3′ LTR sequences were successfully amplified and sequenced from the culture supernatants of three individuals (Pt 7, Pt 8, and Pt 10) (Fig. 1).

The homogeneity in lengths of the 5′ LTRs was also confirmed by sequence analyses. A total of 29 LTR DNA and 9 LTR RNA sequences were obtained. No intrasample length polymorphism was observed in any of the LTS, although a limited degree of intersample length variations was found upstream of the NF-κB binding sites (Fig. 1). When compared to the HIVHXB LTR, Pt 4, Pt 7, and Pt 10 exhibited a 22- to 26-nucleotide insertion. Part of these inserted nucleotides was, in fact, a duplication of the upstream regions, previously designated as TCF-1α sites (8, 10) (Fig. 1). Duplication of TCF-1α in the HIV-1 LTR is not an uncommon finding, as similar observations have been made in both asymptomatic and symptomatic patients (7, 24, 34). Similar patterns of TCF-1α duplication were also found in 3′ LTR RNA sequences obtained from tissue culture supernatants of Pt 7 and Pt 10. Interestingly, no sequence variation was found in 3′ LTR RNA sequences within each individual (Pt 7, Pt 8, and Pt 10) and obtained sequences represent only a minor portion of corresponding proviral LTR DNA sequences (Fig. 1).

### TABLE 1. Characteristics of LTS of HIV-1 infection

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (yr)/sex</th>
<th>Route of infection</th>
<th>Duration of infection (yr)</th>
<th>CD4 cell count (no./mm³)</th>
<th>Isolation of HIV-1 from:</th>
<th>Plasma viral load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt 2</td>
<td>38/M</td>
<td>IV drug use</td>
<td>≥12</td>
<td>500–700</td>
<td>PBMC</td>
<td>&lt;630</td>
</tr>
<tr>
<td>Pt 3</td>
<td>46/M</td>
<td>Homosexual</td>
<td>≥13</td>
<td>560–740</td>
<td></td>
<td>&lt;630</td>
</tr>
<tr>
<td>Pt 4</td>
<td>40/M</td>
<td>Homosexual</td>
<td>14</td>
<td>500–1,200</td>
<td></td>
<td>6,393</td>
</tr>
<tr>
<td>Pt 5</td>
<td>41/M</td>
<td>IV drug use</td>
<td>≥12</td>
<td>800–1,000</td>
<td></td>
<td>&lt;630</td>
</tr>
<tr>
<td>Pt 6</td>
<td>38/M</td>
<td>Homosexual</td>
<td>12</td>
<td>560–860</td>
<td></td>
<td>5,180</td>
</tr>
<tr>
<td>Pt 7</td>
<td>42/F</td>
<td>Heterosexual</td>
<td>13</td>
<td>500–850</td>
<td>+ e</td>
<td>&lt;630</td>
</tr>
<tr>
<td>Pt 8</td>
<td>44/M</td>
<td>Homosexual</td>
<td>≥15</td>
<td>400–800</td>
<td>+</td>
<td>839</td>
</tr>
<tr>
<td>Pt 10</td>
<td>47/M</td>
<td>Homosexual</td>
<td>≥14</td>
<td>550–850</td>
<td>+</td>
<td>11,549</td>
</tr>
</tbody>
</table>

* This table has been presented previously elsewhere (16), and is included here solely for clarity, with permission of the publisher.

* By CD8 depletion only.

* RNA copies/milliliter measured by bDNA assay (Chiron).

* M, male; F, female.

* IV, intravenous.
Some degree of sequence variation was found within previously identified binding sites for both NF-κB and Sp1 (Fig. 1). When compared to the HIVHXB LTR, sequences from Pt 5, Pt 6, Pt 8, and Pt 10 bear multiple mutations in the NF-κB binding sites. A single-base deletion in the first NF-κB binding site (NF-κB I) was found in Pt 4 (Fig. 1). Variations in the Sp1 binding sites were also frequent in all subjects (Fig. 1). However, the presence of G-to-A hypermutations within both the
NF-κB and Sp1 binding sites is most striking in Pt 5. Such hypermutations are expected to have a significant impact on the function of the LTR since they are within the critical NF-κB and Sp1 binding motifs (2, 10, 11) (see below). The degree of sequence variation in the TATA box and the TAR element is much less compared to that in the NF-κB and Sp1 binding sites. Two of the three clones from Pt 4 shared an A-to-T transition in the TATA box. Sequences in Pt 5, Pt 8, and Pt 10 also displayed mutations in the bulge or loop region of the TAR element (Fig. 1). Again, hypermutations, which may disrupt the proper function of the LTR, were found in the TAR element of sequences obtained from Pt 5. This high degree of G-to-A mutations was unlikely the result of errors introduced during the process of PCR amplification, since sequencing the PCR products directly from another PBMC sample revealed identical results (data not shown).

To delineate the genetic features of the 5′ LTR sequences from our cohort of LTS, the degree of sequence diversity was also quantitatively analyzed and compared to sequences (HIVMN, HIVJRCSF, HIVJRFL, HVCAM1, HVSF2, HIVYU2, and HIVHXB) from patients with AIDS, using Kimura’s two-parameter model (22). The degree of intersubject sequence diversity within each LTS varied substantially from case to case, ranging from 0.89% in Pt 4 to 3.49% in Pt 8. The mean intersubject diversity among the LTS (7.40%), however, is not significantly different from that determined from AIDS patients (5.58%) (P > 0.01). Therefore, the degree of sequence variation in the LTR is not associated with the rate of disease progression, a conclusion consistent with previous reports (7, 16, 26, 32, 44).

Analysis of transcriptional activity of the LTR by the luciferase reporter gene system. The impact of the observed mutations in the LTS LTRs on the level of transcriptional activities was evaluated in the context of a luciferase reporter gene system (43). Cloned LTR fragments (Pt 2-4, Pt 3-13, Pt 4-1, Pt 5-5, Pt 6-1, Pt 7-3, Pt 8-1, and Pt 10-3) were subcloned into the upstream region of the luciferase reporter gene in plasmid pGL-2 (Promega) (43). The LTR sequence from HIVHXB was also cloned into the same reporter vector for comparison. The results of transient transfections of these constructs into 293 cells are shown in Fig. 2a.

It is evident that the different LTRs displayed various levels of basal as well as Tat-induced transcription. At basal levels of transcription, the LTS LTR sequences from seven (Pt 2, Pt 3, Pt 4, Pt 6, Pt 7, Pt 8, and Pt 10) of the eight patients yielded expression levels indistinguishable from that of the positive control (HIVHXB LTR) (Fig. 2a). However, the most remarkable finding is the severely blunted activities of the LTR sequences from Pt 5 (Fig. 2a). The high degree of G-to-A mutations in these LTR sequences presents a likely explanation, since such mutations are expected to disrupt the NF-κB and Sp1 consensus binding sites as well as to destabilize the secondary structure of the TAR element (2, 8–10, 12–14, 39).

A wide range of Tat-induced transcriptional activities was found among the LTR sequences from LTS (Fig. 2a). The LTR sequences from Pt 2 displayed the highest level of activity, although no discernible differences could be detected from those of Pt 3, Pt 6, Pt 7, Pt 10, and the positive control HIVHXB LTR (P > 0.01) (Fig. 2a). Significantly lower levels (P < 0.05) of activities were found for LTR sequences from Pt 4 and Pt 8. However, the genetic basis for the reduced level of transcription is currently unclear. As expected, LTR sequences from Pt 5 exhibited marginal levels of transcription that were almost 100-fold less than that of the positive control HIVHXB LTR (Fig. 2a). Furthermore, the transcriptional level of Pt 5's LTR in the presence of Tat is below the basal activities of HIVHXB LTR or of LTR sequences from the other LTS such as Pt 2 or Pt 3. This highlights the effect of G-to-A hypermutations on LTR functional abnormalities (Fig. 1). In multiple, independent experiments, these findings remained consistent.

Analysis of transcriptional activity of the LTR in the context of the whole infectious virus. LTR sequences (Pt 2-4, Pt 3-13, Pt 4-1, Pt 5-5, Pt 6-1, Pt 7-3, Pt 8-1, and Pt 10-3) were also introduced into the infectious provirus HIVHXB-LTRuc to assess the effect of the observed mutations on the level of viral transcription. We replaced the 5′ LTR of HIVHXB-LTRuc with each LTS's 5′ LTR in light of the unique process by which the new 5′ LTR forms upon subsequent infection (43). Twenty micrograms of each proviral construct was used to transfect 293 cells. Virus concentration in the culture supernatant was standardized by measuring the p24 concentration (Abbott Laboratories) prior to use in the infection assay. Five million phytohemagglutinin-stimulated HIV-1-negative donor PBMC were inoculated with 5 ng of HIV-1 p24prov. Luciferase activities
in infected PBMC measured at postinfection are summarized in Fig. 2b.

Figure 2b shows the temporal level of luciferase activity detected in PBMC on days 0, 2, 3, 5, 7, and 12. For all proviral constructs except HIVPt5LTRLuc, peak levels of transcription were recorded approximately 7 days postinfection, with a moderate decline or increase in the ensuing period (Fig. 2b). In the context of the whole infectious virus, the transcriptional activities of LTR sequences from Pt 2, Pt 3, Pt 6, Pt 7, and Pt 10 were comparable to that of HIVHXBLTRLuc ($P > 0.01$). Significantly lower levels ($P < 0.05$) of viral transcription were found under the direction of LTR sequences from Pt 4 and Pt 8, which was remarkably consistent with results obtained using the luciferase reporter gene alone (see above). No detectable level of viral transcription was found in PBMC infected by HIVPt5LTRLuc (Fig. 2b), despite using twice as much virus in the infection (data not shown). The lack of transcriptional activity by Pt 5’s LTR in both assay systems shows the detrimental effect of G-to-A hypermutations on LTR function.

In the present study, we chose to investigate the genotypic and phenotypic properties of the HIV-1 5′ LTR region from eight previously identified LTS. The genetic and functional analyses of nucleotide sequences amplified directly from the PBMC of these LTS revealed a prominent degree of sequence diversity associated with a wide range of basal and Tat-induced transcriptional activity. The observed sequence variations included both point mutations as well as length polymorphisms. Most of the length polymorphisms were associated with a duplication of the TCF-1α fragment, although this duplication is neither complete nor perfect in any of these cases (except Pt 4). Functionally intact 5′LTRs, in terms of directing both heterologous and viral gene expression, were found in seven of the eight LTS, suggesting that genetic and functional defects in the 5′ LTR are unlikely to be the underlying explanation for the well-being of these LTS. However, more thorough future studies are needed to evaluate the biological impact of the observed mutations and length polymorphisms on the level of LTR transcriptional activity.

The LTR sequences from Pt 5, however, stood out both in the genetic and transcriptional studies. A high degree of G-to-A mutations was found in multiple samples throughout the regions important for both NF-kB and Sp1 binding and viral trans-activator Tat function. Functional studies further demonstrated the deficiency of Pt 5’s LTR in directing gene transcription within the context of both the luciferase reporter gene and the whole infectious virus. In our previous study, no detectable level of viral RNA could be found in the plasma samples from Pt 5 (4). Attempts to culture infectious virus from both PBMC and plasma samples were uniformly unsuccessful, raising the possibility that viruses in Pt 5 may be somewhat attenuated (4). In addition, Pt 5 is able to mount vigorous humoral and cytotoxic immune responses against a diverse panel of laboratory-adapted as well as primary HIV-1 isolates (4). A robust immune response may, therefore, have shifted the balance toward the host in the face of Pt 5’s sequentially and functionally defective LTR, providing some explanations for his well-being over a prolonged period of time.

This work was supported by grants from the National Institutes of Health (AI35168, AI45218, and AI40387) and the Aaron Diamond Foundation.

REFERENCES