

# Molecular Characterization of Long Terminal Repeat Sequences from Brazilian Human Immunodeficiency Virus Type 1 Isolates

Geraldo A. Ferraro,<sup>1,2</sup> Joana P. Monteiro-Cunha,<sup>1-3</sup> Flora M.C. Fernandes,<sup>1-3</sup> Aline C.A. Mota-Miranda,<sup>1-3</sup> Carlos Brites,<sup>3</sup> Luiz C.J. Alcantara,<sup>1,2</sup> Bernardo Galvão-Castro,<sup>1,2</sup> and Mariza G. Morgado<sup>4</sup>

## Abstract

HIV-1 provirus activation is under control of the long terminal repeat (LTR)-5' viral promoter region, which presents remarkable genetic variation among HIV-1 subtypes. It is possible that molecular features of the LTR contribute to the unusual profile of the subtype C epidemic in the Brazilian Southern region. To characterize the LTR of Brazilian HIV isolates, we analyzed sequences from 21 infected individuals from Porto Alegre and Salvador cities. Sequences were compared with subtype B and C reference strains from different countries. Phylogenetic analysis showed that 17 (81%) samples were subtype B and four (19%) were subtype C. Common patterns of transcription factor binding sites (TFBS) in subtypes B and C sequences were confirmed and other potential TFBS specific for subtype C were found. Brazilian subtype C sequences contained an additional NF- $\kappa$ B binding site, as previously described for the majority of subtype C isolates. The high level of LTR polymorphisms identified in this study might be important for viral fitness.

## Introduction

THE HIV-1 5' LONG TERMINAL REPEAT (LTR-5') is related to transport, integration, and control of virus transcription. The general structure of the HIV-1 LTR is subdivided into three regions: U3, R, and U5. The U3 region harbors important transcription factor binding sites (TFBS) that are related to the promoter, enhancer, and modulatory regions.<sup>1</sup> The core of the promoter and enhancer regions contains binding motifs for Sp1 (Specific protein 1) and nuclear factor kappa B (NF- $\kappa$ B), respectively. Both of these are essential for Tat activity. In addition to the Tat protein, which binds to the TAR hairpin on the R region, the host cell transcription factors (TF) are also involved in the control of HIV-1 transactivation.<sup>1</sup> The enormous level of HIV-1 polymorphism represents an obstacle to drug development and can, additionally, be related to HIV-1 pathogenesis.<sup>2</sup> The diversity of clinical outcome patterns may be related to these numerous variants. HIV-1 tropism specificity is primarily dictated by the CD4 protein and the che-

mokine receptors expressed on the cell surface.<sup>3</sup> Thus, HIV *env* glycoproteins play a fundamental role in viral entry.

The intracellular environment is, however, the main determinant of viral fitness. In fact, specific TF sets in each cell type and the respective binding sequences available within the LTR promoter region influence viral replication and fitness.<sup>4</sup> The genetic variation of the LTR is probably under the control of the same processes that drive the diversity of other HIV-1 genetic regions. The rapid turnover of viral particles, the lack of DNA repair mechanisms during the reverse transcription process, and recombination events play an important role in the generation of viral diversity.<sup>5</sup> The HIV-1 LTR is especially variable and this feature may determine distinct patterns of virus replication. Several studies have reported the presence of cis-acting elements (Sp1, NF- $\kappa$ B, c-Rel, C/EBP, NF-AT, Ets-1, GATA) and their respective functional trans-regulatory proteins for viral transactivation.<sup>1</sup> In the present work, molecular features of HIV-1 subtypes B and C LTR-5' obtained from 21 Brazilian isolates

<sup>1</sup>Laboratório Avançado de Saúde Pública (LASP), Centro de Pesquisa Gonçalo Moniz (CPqGM), Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil.

<sup>2</sup>Fundação Bahiana para o Desenvolvimento das Ciências (FBDC), Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador, Bahia, Brazil.

<sup>3</sup>Universidade Federal da Bahia, Salvador, Bahia, Brazil.

<sup>4</sup>Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz/FIOCRUZ, Rio de Janeiro, RJ, Brazil.

were analyzed. The major sites of the LTR-5' fragment were described and potential transcription factor binding sites were identified.

## Materials and Methods

### Study population

Blood samples were obtained in 2003 from 16 HIV-1-infected individuals, regularly followed up at the Hospital de Clínicas de Porto Alegre, Rio Grande do Sul, Brazil and from five patients followed up at Hospital Universitário Edgard Santos, Salvador, Bahia, Brazil. Subjects signed a written consent form for their inclusion in the study. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque (Sigma Chemical Co., St. Louis, MO) density gradient centrifugation for viral isolation and DNA extraction. The Ethical Boards of both hospitals approved this work.

### DNA isolation, polymerase chain reaction (PCR), and sequencing

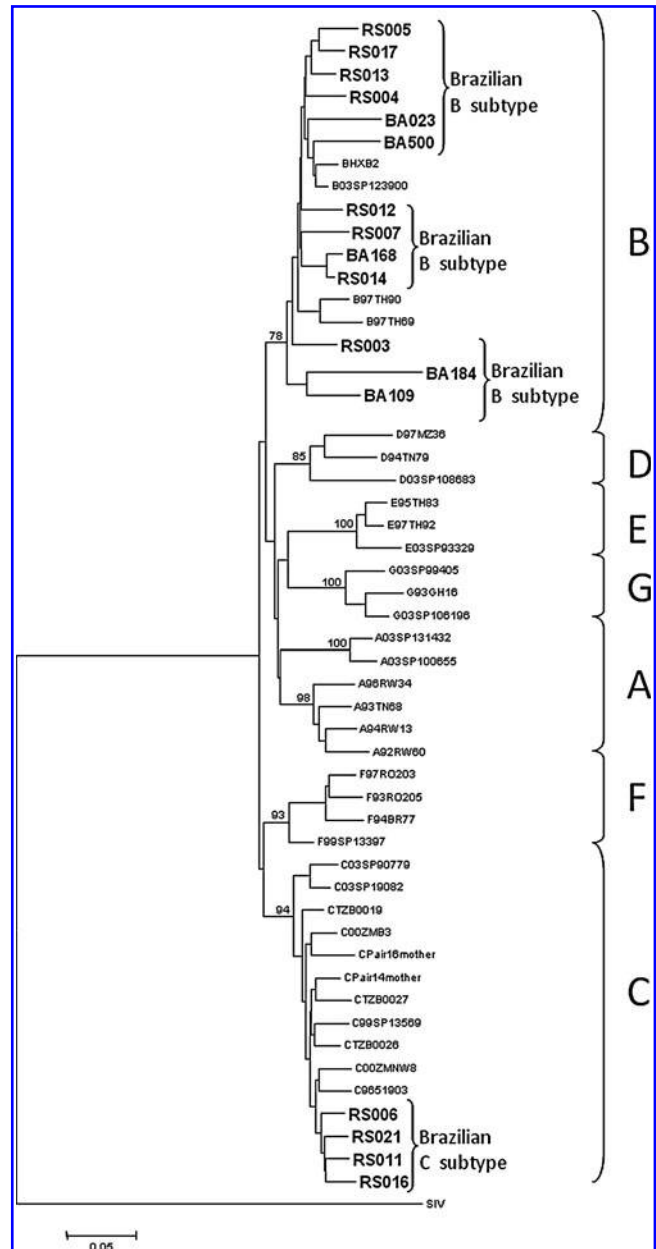
DNA was extracted using the Qiagen extraction kit (QIAGEN, Valencia, CA). Nested PCR was conducted for individual amplification of LTR-5' (~690 bp) in a Perkin Elmer 9600 Thermal Cycler. The LTR-5' amplification was performed using LTR1-F/JL17-R and LTR2-R/LTR3-F as the outer and inner primer sets, respectively.<sup>6</sup> PCR products were purified in Qiagen columns (QIAGEN, Valencia, CA) and sequenced in an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) using a Big Dye Terminator kit (Applied Biosystems, Foster City, CA) and the same PCR inner primers. Sequences were assembled using the SeqMan software (DNASar, Madison, WI).

### Phylogenetic analysis

LTR sequences from this study ( $n=21$ ) and 35 reference strains from the GenBank/EMBL database were aligned with Clustal X<sup>7</sup> and edited manually using GeneDoc.<sup>8</sup> Neighbor-joining (NJ) tree reconstruction was performed in the Mega 4.0 package<sup>9</sup> with the Kimura two-parameter model of nucleotide substitution. The reliability of the NJ trees was assessed by analyzing 1,000 bootstrap pseudoreplicates.

### Characterization of potential TFBS

To identify the potential TFBS, 21 new LTR sequences (17 subtype B and 4 subtype C) from Brazilian viruses plus 15 non-Brazilian sequences (4 subtype B and 11 subtype C) were selected. The non-Brazilian sequences were downloaded from GenBank. The selection criterion for these isolates was based on the availability of phylogeny data and adequate sequence length. The data set was submitted to the MatInspector<sup>10</sup> online bioinformatics tool after alignment and edition. These analyses allowed for the identification of potential TFBS within distinct cell lines. The samples were analyzed in two groups according to the presence of potential TFBS activity in (1) nonneuroglia and (2) neuroglia cells. Other additional approaches to identify specific features on the LTR were carried out throughout the alignment of the motifs related to TFBS as the Sp1, NF- $\kappa$ B, and TAR region.

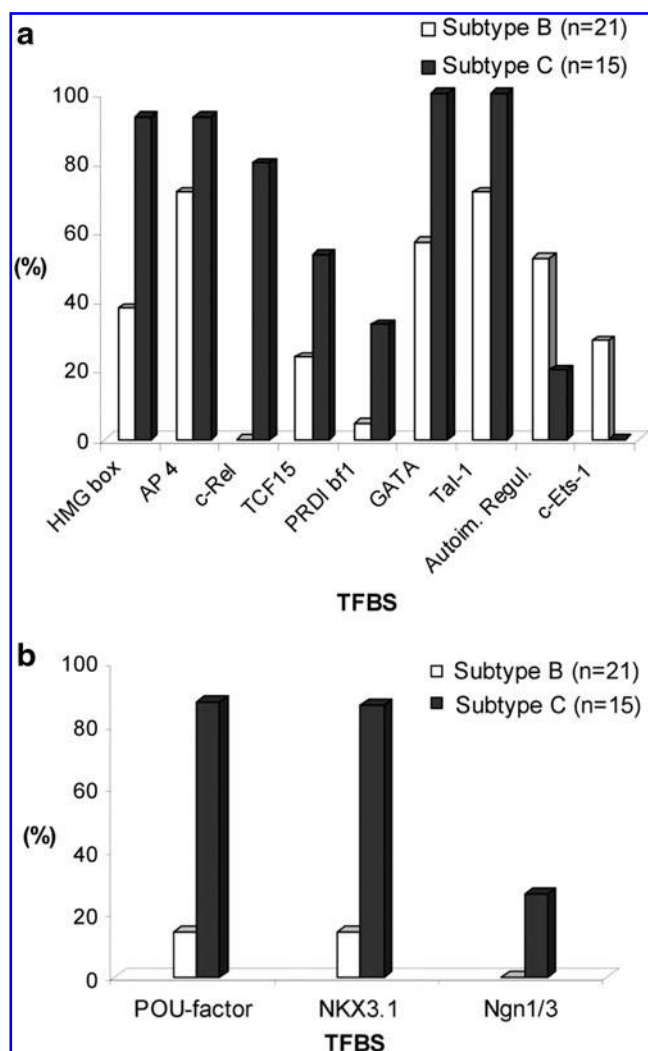


**FIG. 1.** Rooted neighbor-joining tree based on long terminal repeat (LTR) sequences (450 bp) of Brazilian HIV-1 isolates. HIV-1 subtype reference sequences were obtained from the GenBank/EMBL databases. The tree was rooted using an SIV sequence (accession M29973). Bootstrap values >70% were considered significant. Newly sequenced LTR sequences of Brazilian HIV-1 isolates included in this analysis are in bold.

## Results

### Phylogenetic analysis of LTR-5' sequences

The final length of the LTR dataset after edition comprised 450 bp. In the phylogenetic analysis (Fig. 1), the majority of the HIV-1 isolates (17 out of 21) clustered within the subtype B clade (bootstrap = 78%). It is worth noting that four subtype B samples are not shown in the tree, because they had shorter LTR sequences; these sequences were analyzed through



**FIG. 2.** Potential transcription factor binding sites (TFBS) with activity in nonneuroglia (a) and neuroglia (b) cells, identified in the LTR-5' region of subtype B and C HIV-1 isolates. The cis elements were identified by the MatInspector bioinformatics tool. The TF binding is shown as percent value for each HIV-1 subtype.

individual phylogenetic analysis (not shown). The remaining four isolates (RS006, RS021, RS016, and RS011) clustered into the subtype C clade (bootstrap = 94%). All subtype C isolates were from Porto Alegre city (25%). On the other hand, all viruses from Bahia were typed as subtype B.

#### Characterization of potential and regular transcription factor binding sites and TAR region

The results of the analyses of potential LTR binding sites specific to transcriptional factors with activity in neuroglia and nonneuroglia cell types are depicted in Fig. 2. The MatInspector analysis showed a higher incidence of potential TFBS within the LTR of subtype C when compared to subtype B (HMG-box, c-Rel, NKX3.1, Pou-factor, and ngn1/3). Several potential TFBS with activity in nonneuroglia cells showed little or no difference in their frequencies between subtypes B and C. However, differences could be observed in some potential TFBS frequencies between subtypes: HMG-box (–61 to

–75 nt relative to HXB2; 94% for subtype C and 25% for subtype B) and c-Rel (–93 to –80 nt relative to HXB2; 81% for subtype C and 0% for subtype B).

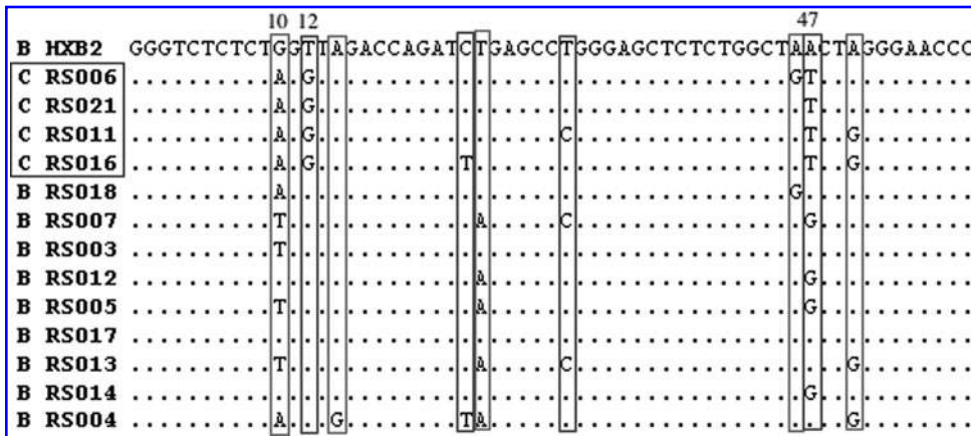
The analysis of the potential TFBS with activity in neuroglia cells showed that 94% of the subtype C isolates presented at least one potential binding site to Pou-factor and 67% of them had a double Pou-factor binding site (–207 to –192 nt and –142 to –130 nt relative to HXB2). The potential TFBS to Pou-factor was present in 15% of the subtype B isolates. The presence of potential TFBS to NKX3.1 (–171 to –155 nt relative to HXB2) was also found in subtype C (88%) and subtype B (15%) isolates. The presence of the potential ngn1/3 TFBS (108 to 120 nt relative to HXB2) was exclusive to subtype C (25%). All isolates of subtypes B or C contained at least one X-Box potential binding site at position 43 to 61 nt and 121 to 138 nt relative to the HXB2 reference sequence. Those spanning positions 43 to 61 were less conserved and were usually not recognized as x-Box sites by the MatInspector software. This occurred especially with subtype C sequences. The evaluation of Sp1 tandem binding sites (III, II, and I) downstream of the NF- $\kappa$ B binding sites (–81 to –46 nt relative to HXB2) was done in all new Brazilian and reference sequences.

Two specific nucleotide signatures were identified in the Sp1(III) of subtype C sequences (A at the fourth position and T at the ninth position). The Sp1(II) and (I) sites were more conserved than Sp1(III). All isolates presented the TATA-box binding site in the promoter region with the common sequence CATATAAG at position –28 to –24 nt relative to HXB2 (not shown). Four isolates of subtype B from Porto Alegre presented an additional TATA-box or a TATA-box-like, as a potential TFBS downstream from the first one (153 to 160 nt relative to HXB2). Two of them had a sequence CATATAGT, one CATTATAGT with one T insertion, and the last one had GCTATAAG. All isolates contained the canonical NF- $\kappa$ B double site (spanning –109 to –84 nt relative to HXB2) with the GGGACTTCC sequence. None of the NF- $\kappa$ B binding sites II and I had the conserved GGGA core followed by the CTTTCC subjacent sequence in subtype B viruses. The Brazilian subtype C viruses contained a deletion of an A (GGG–) and a substitution of a T to a G (CGTTC) at the fourth and sixth position, respectively, within the second Sp1 tandem site. The additional NF- $\kappa$ B binding site (–119 to –113 nt relative to HXB2) was identified in almost all subtype C isolates, including the Brazilian strains. Only one subtype C isolate did not present the third NF- $\kappa$ B binding site (92br025). The TAR region alignment showed polymorphic sites in the B and C isolates from Porto Alegre (Fig. 3).

At least three subtype C-specific mutations were identified: a G to A substitution at the tenth position, a T to G substitution at the twelfth position, and a A to T substitution at position 47. In contrast, a A to G substitution was predominant at position 47 among the subtype B isolates. The secondary structure of this region in both subtypes lacked some hydrogen bonds (data not shown).

#### Discussion

The high polymorphism of HIV-1, especially in the structural genes, and its influence on the HIV/AIDS epidemic prompted us to characterize the molecular features of the LTR region in HIV-1 subtypes in Brazilian patients. Phylogenetic analyses based on the LTR region showed 25% of subtype C viruses in Porto Alegre city. These data compared with recent



**FIG. 3.** Multiple alignment of TAR sequences from HIV-1 subtypes B and C isolates from Porto Alegre city, showing polymorphic sites. Amino acid positions above the alignment are numbered according to their positions in the TAR region.

molecular epidemiological studies based on the *pol* gene, showing 56.3% of subtype C, provides evidence for the increase in subtype C prevalence in Southern Brazil in the past decade.<sup>11</sup> Although subtype F was not identified among our LTR sequences, additional analysis (unpublished data) showed that five (23.8%) isolates (BA023, BA109, BA184, BA168, and BA500) were classified as subtype F in the *gag* region, whereas they were typed as B in the LTR region. At least two other studies reported B/F recombinants from Brazil whose LTR regions were also B,<sup>12,13</sup> suggesting that the subtype F LTR could be less efficient than the subtype B LTR.

Viral genetic diversity is related to the ability of virus to replicate. The maintenance of HIV-1 genetic variability is an important strategy to facilitate fitness and to colonize new cell types and spread the infection to different tissues. The different frequencies of potential TFBS determine typical subtype-specific patterns for subtype B and C viruses. This could in turn be related to HIV-1 gene regulation control and, consequently, to the AIDS epidemic profile in regions in which these subtypes predominate.

The higher frequency of potential TFBS with activity in neurological cells supports the hypothesis that the subtype C LTR region could be related to productive infection in the central nervous system. This might lead to severe clinical outcomes including dementia in AIDS patients. In addition, the importance of the host environment to virus fitness has been demonstrated in different cell lines.<sup>14</sup> The predominance and frequency of specific potential TFBS with activity in neuroglia cells in subtype C sequences are strong evidence for their importance to viral fitness. Further studies assembling clinical and molecular data will be necessary to clarify this point. The four Brazilian subtype C samples from Porto Alegre contained three NF- $\kappa$ B binding sites. This common molecular feature is found in the majority of subtype C described in the global epidemic. Curiously, one subtype C Brazilian sample (92BR025) did not contain the extra NF- $\kappa$ B binding site, which could imply multiple introduction events for subtype C in Brazil.

The role of the Sp1 triple site on HIV-1 gene activation has been well described.<sup>15</sup> The presence of this cis element is critical to virus viability and a recent study demonstrated the relationship between the genetic variability of these sites and clinical features of HIV-1-infected patients.<sup>2</sup> The Sp1 (III, II, I) nucleotide sequences were well conserved in at least half of

the Brazilian HIV-1 isolates. The Sp1 (II) and (I) sites were further conserved and a peculiar signature was identified in the Sp1 (III) binding site of Brazilian subtype C samples. These conserved sites might be related to the improvement in the viral transactivation mechanism and might be linked to typical HIV progressor patients.<sup>2</sup> In fact, the majority of the patients from Porto Alegre presented opportunistic diseases during clinical follow-up and some of them had already died.

The TAR region analysis showed that Brazilian strains presented some polymorphic sites as did the non-Brazilian isolates, including at least two subtype-specific mutations related to B or C viruses (Fig. 3). The study of the hairpin RNAs using the secondary structure model showed no drastic change in their architecture (data not shown). However, some mutations leading to the absence of hydrogen bonds between the antiparallel strands suggest that a more flexible structure could modulate Tat binding.

In conclusion, we present a systematic characterization of the LTR-5' region from Brazilian HIV-1 subtypes B and C. Some typical subtype-specific potential TFBS were identified. The importance of Sp1, NF- $\kappa$ B, and TATA-box motifs as well as the TAR region to virus viability was confirmed. An additional NF- $\kappa$ B site among Brazilian subtype C isolates was identified. Considering the high variability within the LTR region presented in this study, additional studies associating functional and molecular information should be conducted to elucidate the role of these LTR polymorphisms.

### Sequence Data

The sequences were reported to the GenBank database under accession numbers JX965917 to JX965937.

### Acknowledgments

The authors are grateful to Luciano Zulbaran Goldane for the selection of samples from Porto Alegre and David Watkins, visiting professor (CAPES-FIOCRUZ), for revising this manuscript. This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), the Brazilian Ministry of Health STD/AIDS Program, CNPq, and FIOCRUZ.

### Author Disclosure Statement

No competing financial interests exist.

## References

1. El Kharroubi A and Martin M: Cis-acting sequences located downstream of the human immunodeficiency virus type 1 promoter affect its chromatin structure and transcriptional activity. *Mol Cell Biol* 1996;16:2958–2966.
2. Ramírez EA, Martín C, Soriano V, *et al.*: Genetic analysis of the long terminal repeat (LTR) promoter region in HIV-1-infected individuals with different rates of disease progression. *Virus Genes* 2007;34:111–116.
3. Berger EA, Murphy PM, and Farber JM: Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu Rev Immunol* 1999;17:657–700.
4. Ranjbar S, Tsytsykova AV, Lee S-K, *et al.*: NFAT5 regulates HIV-1 in primary monocytes via a highly conserved long terminal repeat site. *PLoS Pathogens* 2006;12:1176–1186.
5. Preston BD, Poiesz BJ, and Loeb LA: Fidelity of HIV-1 reverse transcriptase. *Science* 1988;242:1168–1171.
6. Varas F, Medrano L, Ballester S, and Nájera R: Influence of PCR parameters on amplifications of HIV-1 DNA: Establishment of limiting sensitivity. *Biotec* 1991;11:384–386, 388–391.
7. Jeanmougin F, Thompson JD, Gouy M, *et al.*: Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998;23:403–405.
8. Nicholas KB, Nicholas HB Jr, and Deerfield DW II: GeneDoc: Analysis and visualization of genetic variation. *EMBNET News* 1997;4:1–4.
9. Tamura K, Dudley J, Nei M, and Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599.
10. Cartharius K, Frech K, Grote K, *et al.*: MatInspector and beyond: Promoter analysis based on transcription factor binding sites. *Bioinformatics* 2005;21:2933–2942.
11. Silveira J, Santos AF, Martínez AM, *et al.*: Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J Clin Virol* 2012;54:36–41.
12. De Sa Filho DJ, Sucupira MC, Caseiro MM, *et al.*: Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res Hum Retroviruses* 2006;22:1–13.
13. Sanabani S, Neto WK, de Sa Filho DJ, *et al.*: Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res Hum Retroviruses* 2006;2:171–176.
14. van Opijnen T, de Ronde A, Boerlijst MC, and Berkhout B: Adaptation of HIV-1 depends on the host-cell environment. *PLoS ONE* 2007;2:e271.
15. Yedavalli VS, Benkirane M, and Jeang KT: Tat and transactivation-responsive (TAR) RNA-independent induction of HIV-1 long terminal repeat by human and murine cyclin T1 requires Sp1. *J Biol Chem* 2003;278:6404–6410.

Address correspondence to:

Joana P. Monteiro-Cunha  
Departamento de Biofunção  
Instituto de Ciências da Saúde  
Universidade Federal da Bahia  
Av. Reitor Miguel Calmon s/n-Vale do Canela  
Salvador, Bahia 40110-100  
Brazil

E-mail: joanapmonteiro@hotmail.com