

The Human Immunodeficiency Virus Type 1 Promoter Contains a CATA Box Instead of a TATA Box for Optimal Transcription and Replication

Tim van Opijnen,^{1,2} Joost Kamoschinski,¹ Rienk E. Jeeninga,¹ and Ben Berkhout^{1*}

Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam,¹ and Section of Population Biology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1098 SM Amsterdam,² The Netherlands

Received 12 December 2003/Accepted 18 February 2004

The human immunodeficiency virus type 1 (HIV-1) transcriptional promoter contains a single polymorphism in the TATA box. Most subtypes contain the sequence TATAAGC, but subtype E and some recombinant AG strains have the sequence TAAAAGC. Based on mutagenesis studies of cellular RNA polymerase II (pol II) promoters, it has been proposed that the subtype E TATA box is nonfunctional due to the T-to-A substitution at the critical position 3. By means of transcription and virus replication assays, we demonstrate that the true TATA box motif within the viral long terminal repeat (LTR) promoter starts two nucleotides further upstream. Because of this realignment, subtype E has the sequence CATAAAA and all other subtypes have the sequence CATATAA. The polymorphism therefore has shifted from position 3 to position 5 and is no longer incompatible with efficient transcription according to rules determined for cellular pol II promoters. In addition, through sensitive competition experiments, we demonstrate that the CATA box of subtypes B and E can be improved for replication by the mutations 1T and 5T, respectively. The fact that the fitness of both subtype LTRs can be increased by specific point mutations in the CATA box suggests that the transcriptional promoter of HIV-1 is fine-tuned towards a suboptimal level of replication. However, this replication rate may be optimal in the in vivo context of an infected individual.

The current AIDS pandemic is caused by at least nine human immunodeficiency virus type 1 (HIV-1) subtypes (termed A through K) and an increasing number of recombinant forms. The clustered dispersal of the HIV-1 subtypes around the world raises questions concerning the determinants that have shaped this pattern. A possibility is that some subtypes spread faster due to an elevated replication rate or more efficient transmission (5, 8, 10, 11, 22, 23). Alternatively, the current geographical distribution might have originated due to chance. Recently, we and others have shown that the viral long terminal repeat (LTR) contains subtype-specific transcription factor binding sites (TFBS), which influence the replication rate in a subtype-specific manner (9, 15, 26). Moreover, the exact strength of the promoter depends on the interaction between the TFBS in the LTR and the pool of active transcription factors in the cellular environment (26). The LTR can be subdivided into the U3, R, and U5 regions, the first two of which are essential for transcription. Depending on the HIV-1 subtype, different combinations of TFBS, for instance, ETS-1, AP-1, NF- κ B, and Sp-1, are found in the U3 region. Upon transcription, the R region of the nascent transcript folds the TAR hairpin, which interacts with the viral transactivator protein Tat to fully activate transcription. The HIV-1 TATA box is located 28 bp upstream of the transcription start site in the U3 region (4, 20). This important transcriptional motif comes in two variants; most subtypes have the sequence TATAAGC,

but all subtype E isolates and some recombinant AG strains (CRF02-AG) have TAAAAGC (Fig. 1). It has been suggested that the subtype E TAAAAGC box may not be functional because the point mutation violates the rules for an active TATA box (7, 30). In this study, we set out to determine whether this variant subtype E TAAAAGC box is functional and whether it can be exchanged with the regular TATAAGC sequence.

The majority of eukaryotic genes, including the proviral genome of HIV-1, encode an RNA polymerase II (pol II) promoter with a TATA box resembling the consensus sequence TATAAAA. The TATA binding protein (TBP) binds in the minor groove of this motif as one of the first steps towards initiation of transcription. TBP is highly conserved across many species; the DNA binding domain is identical in humans and mice, and there is 80% amino acid similarity between human and yeast TBPs (21). There is little free TBP in mammalian cells; it rather is present in complexes with other proteins referred to as TBP-associated factors (1). TFIID is the ~750-kDa TBP-containing complex that interacts with the TATA box and an extensive flanking region of ~70 to 80 bp (21). There are at least seven positions within the TATA box that can have a strong effect on the level of transcription. Mutagenesis of the TATA box in yeast and human model systems has revealed the impact of nonconsensus nucleotides at these positions on the transcriptional activity (7, 18, 30). The highest activity is measured for a TATA box that consists exclusively of T's and A's (TATAAAA). The third position is one of the most critical residues and must be a T (7, 18, 30). This rule is broken by the TATA box of HIV-1 subtype E (TAAAAGC), which suggests that this motif may not be functional. Nevertheless,

* Corresponding author. Mailing address: Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. Phone: 31 20 566 4822. Fax: 31 20 691 6531. E-mail: b.berkhout@amc.uva.nl.

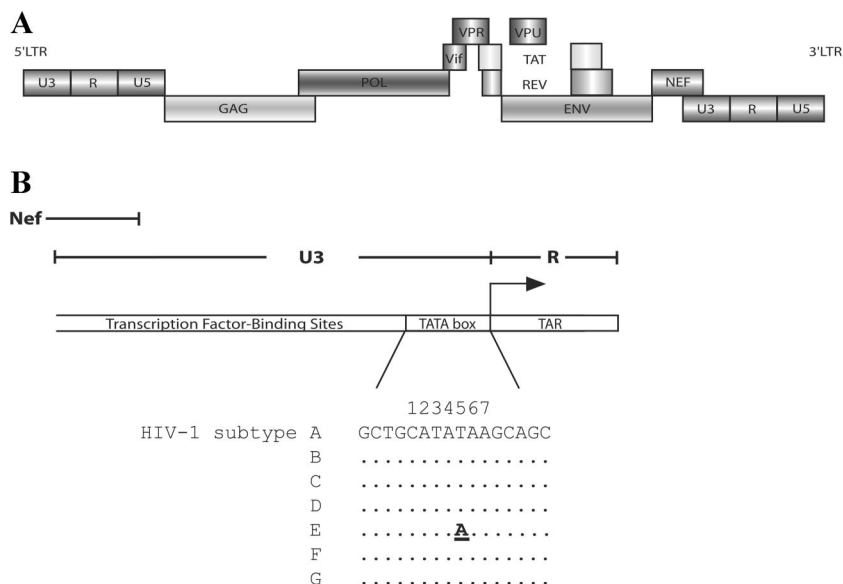


FIG. 1. HIV-1 LTR promoter and CATA boxes of different subtypes. (A) Proviral genome of HIV-1 (not to scale). (B) Part of the LTR promoter and the consensus sequence of the TATA box with flanking nucleotides. The only variation within the TATA box is the T/A polymorphism at position 5 in subtype E versus all other subtypes. The HIV-1 TATA box is known as the TATAAGC motif (positions 3 to 9), but we demonstrate in this study that the motif is in fact CATATAA (positions 1 to 7).

the subtype E LTR is fully active as a transcriptional promoter in transient-transfection assays and virus replication studies (9, 15, 26). Four hypotheses were previously put forward to explain this apparent contradiction (9, 13). First, it was thought that there may be an alternative TATA box elsewhere in the LTR that makes up for the nonconsensus TATA box at position -28 , but this possibility was subsequently ruled out (9). Second, it was thought that the nonconsensus TAAAAGC box may be compensated for by subtype E-specific mutations in the TAR hairpin (13); however, this explanation was contradicted by sequence data (16) and experimental results (9). Third, subtype E may use the nonconsensus TAAAAGC box due to compensatory mutations in other TFBS. Fourth, transcription in subtype E may be regulated by a TATA-less promoter (17).

We tested two sets of TATA box mutations in transient LTR-luciferase transfection assays and in virus replication studies. The results unequivocally demonstrate that HIV-1 does not use the previously defined TATA box (TATAAGC) but instead uses the motif that starts two nucleotides upstream: the CATA box (CATATAA). In this realigned sequence context, the subtype E-specific mutation affects position 5 of the CATA box instead of position 3 (CATAAAA). Such sequence variation at position 5 has only a minor effect on the transcriptional activity of the LTR, a result that is consistent with previous mutagenesis studies with other promoters. Thus, the TATA box of all HIV-1 viruses has been misnamed and misplaced and should be a CATA box, a sequence motif that is used by approximately 12% of cellular pol II promoters.

MATERIALS AND METHODS

LTR-luciferase constructs and HIV-1 molecular clones. The pBlue3'LTR-luciferase constructs contain a subtype-specific fragment from either subtype B or E (LTR positions -147 to $+63$). These pBluescript KS(+) derivatives were described previously (9). TATA box mutants were made in the pBlue3'LTR

intermediate plasmid through a PCR strategy with a mutagenic primer (9). Subsequently, the XhoI-BglII fragment with the subtype-specific LTR and the mutated TATA box was inserted into the pBlue3'LTR-luciferase plasmid. Molecular HIV-1 clones with a subtype-specific B and E LTR were described previously (26). Molecular clones with a mutant TATA box were obtained by insertion of the respective XhoI-HindIII fragments of pBlue3'LTR into the 3' LTR of the molecular clone pLAI from subtype B (19). The 3' LTR is inherited in both LTRs of the viral progeny after the first round of virus replication. The R region (including the TAR element) is inherited from the 5' LTR sequence. Therefore, all viruses contain a mutant LTR segment from position -147 to -1 , including the TATA box, but contain the wild-type (wt) TAR element of subtype B.

Cell lines. The human lymphocytic SupT1 T-cell line (2) was cultured in RPMI 1640 (Gibco BRL) supplemented with 10% fetal calf serum, penicillin (100 U/ml), and streptomycin (100 U/ml). The cervical carcinoma cell line C33A and HeLa cells (ATCC HTB31) (24) were cultured in Dulbecco's modified Eagle's medium (Gibco BRL) with the same supplements. All cell lines were kept at 37°C with 5% CO_2 .

Transfection and luciferase assay. C33A and HeLa cells were grown as monolayers and transfected by the calcium phosphate method as described previously (6). To determine the basal transcriptional activities of the different LTR-luciferase constructs, we transfected 20, 40, and 100 ng of plasmid DNA. Mutant LTR activity was calculated relative to that of the wt LTR which, was set at 1 for each experiment. Tat-activated levels of transcription were determined in cotransfections with 20 and 40 ng of the LTR-luciferase construct and 30, 50, and 100 ng of pcDNA3-Tat (27). All measurements were performed in the linear range of LTR transcriptional activity. Transient-transfection assays were performed six to eight times with different preparations of the LTR-luciferase plasmids. Luciferase expression of the first set of mutants was determined 2 days after transfection. Cells were harvested, washed with phosphate-buffered saline and solubilized in reporter lysis buffer (Promega) for 45 min at room temperature. Samples were diluted 1 in 10 in reaction buffer (3.3 mM ATP, 25 mM glycylglycine [pH 7.8], 15 mM MgSO_4 , 100 μg of bovine serum albumin/ml) with 1 mM luciferin (Boehringer Mannheim). For the second set of mutants, we performed a cotransfection with a renilla-encoding plasmid (dual luciferase assay; Promega). Renilla and luciferase activities are measured in the same assay, and the former is used as an internal control for the transfection efficiency. The transcriptional activities of the constructs in the dual luciferase assay are calculated as the ratio of renilla and luciferase activities. When indicated, we cotransfected an expression vector that encodes a mutant TBP (TBP_{AS}) that favors the variant sequence TGTA_{AS} (25). Two amounts were used (25 and 50 ng), which gave

TABLE 1. TATA box mutants for subtypes B and E

Subtype	Set	Mutant	Nucleotide at position ^a :				
			1234567				
B	1	wt	CTG	CATATAA	GCA		
		5A	CTG	CATAAAA	GCA		
		5G	CTG	CATAGAA	GCA		
		5C	CTG	CATACAA	GCA		
		NS	CTG	<u>CTACGAT</u>	GCA		
	2	wt	CTG	CATATAA	GCA		
		1A	CTG	<u>A</u> ATATAA	GCA		
		1T	CTG	<u>T</u> ATATAA	GCA		
		1G	CTG	<u>G</u> ATATAA	GCA		
		2G	CTG	C <u>G</u> TATAA	GCA		
		4G	CTG	CAT <u>G</u> TAA	GCA		
		1T2G	CTG	<u>TGT</u> TATAA	GCA		
		E	1	wt	CTG	CATAAAA	GCA
				5T	CTG	CATA <u>T</u> AA	GCA
5G	CTG			CATAGAA	GCA		
5C	CTG			CATACAA	GCA		
NS	CTG			<u>CTACGAT</u>	GCA		
2	wt		CTG	CATAAAA	GCA		
	1A		CTG	<u>A</u> ATAAAA	GCA		
	1T		CTG	<u>T</u> ATAAAA	GCA		
	1G		CTG	<u>G</u> ATAAAA	GCA		
	2G		CTG	C <u>G</u> TAAAA	GCA		
4G	CTG	CAT <u>G</u> AAA	GCA				
1T2G	CTG	<u>TGT</u> AAAA	GCA				

^a Mutations are underlined.

the same relative increase in transcription activity of the mutants tested. Control cotransfections were performed with a wt TBP expression plasmid (TBP_{wt}).

Virus stocks and virus replication. C33A cells were calcium phosphate transfected with 5 µg of the pLAI molecular clones to produce virus stocks as described previously (6). The virus concentration was determined by the CA-p24 enzyme-linked immunosorbent assay as described previously (9). For virus replication, SupT1 cells (1.25×10^6 /5 ml) were infected with 20 ng of CA-p24 virus stock, and CA-p24 production was monitored over time.

Virus competition experiments, proviral DNA isolation, and sequencing. Competition experiments were conducted with SupT1 T-cells to determine the fitness of TATA box mutants relative to the parental wt virus (26). Pairwise competitions were started at a 1:1 ratio and were performed six times with two independent virus stock preparations. Total cell DNA was isolated from approximately 0.25×10^6 cells in order to sequence the LTRs of the integrated proviruses and to determine the frequency of each competitor. Cells were solubilized in 500 µl of lysis buffer (10 mM Tris [pH 8.0], 1 mM EDTA, 0.5% Tween 20, and 5 ng of proteinase K per ml) for 1 h at 56°C and 10 min at 95°C. Proviral LTR sequences were PCR amplified and sequenced with the T7 DYEnamic direct cycle sequencing kit (Amersham) as described previously (26). We determined the relative pairwise fitness by comparing the initial and final ratios of the viral genotypes (26).

RESULTS

The T/A polymorphism is at position 5 of the TATA box and is compatible with transcription. We tested several TATA box mutations (Table 1) in HIV-1 subtypes B and E in transient-transfection assays for their effect on transcriptional activity. Due to realignment of the TATA box (see below), the first position has been shifted two nucleotides upstream to the C at position -30 (relative to the transcription start site), and this numbering will be used throughout this study. The first set of mutants was designed to elucidate the importance of the

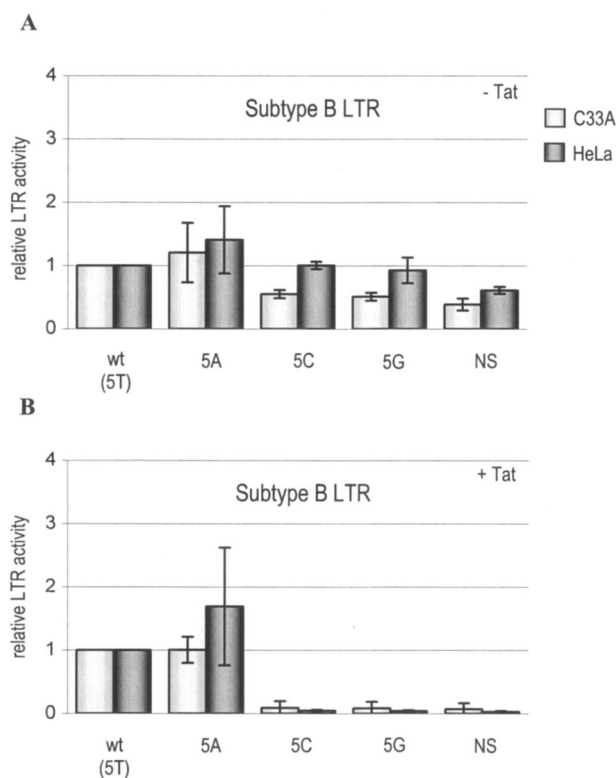


FIG. 2. Transcriptional activities of subtype B TATA box mutants. Shown are the relative luciferase activities of four TATA box mutant LTRs and the wt subtype B LTR. The activity of the wt TATA box sequence was arbitrarily set at 1 in assays without (A) and with (B) the transactivator protein Tat. Transfections were performed eight times in C33A and HeLa cells. Error bars indicate standard deviations.

TATA box for transcription and whether we could exchange the TATA box sequences of subtypes B and E. Position 5 in the TATA box was changed into all three other possible nucleotides (A, G, and C for subtype B and T, G, and C for subtype E). As a negative control, we mutated the TATA box of each subtype into the nonsense sequence CTACGAT (NS mutants). LTR-luciferase transfection assays were performed with two cell types (C33A and HeLa) either with or without the transactivator protein Tat. The results for subtype B and E are shown in Fig. 2 and 3, respectively. TATA box mutations have only a minor effect on the basal level of transcription, which is consistent with previous results obtained with other LTR mutants (3). The transactivator protein Tat induces the wt LTR promoter approximately 100-fold over basal activity, but major defects in Tat-induced LTR activity were observed for some of the TATA mutants. The T/A polymorphism at position 5 is compatible with high-level transcription, but all other mutants (5C, 5G, and NS) were largely inactive. In fact, one could regard the TATA box as a Tat-responsive motif because it is more critical at Tat-induced high levels of transcription. Similar results were obtained in the subtype B and E contexts. This shows that position 5 is important for transcription but that both subtypes accept the T/A sequence variation that naturally occurs in subtype E (100% A) versus all other subtypes (T). The results for the NS mutants indicate that transcriptional activity is negligible when the TATA box is eliminated.

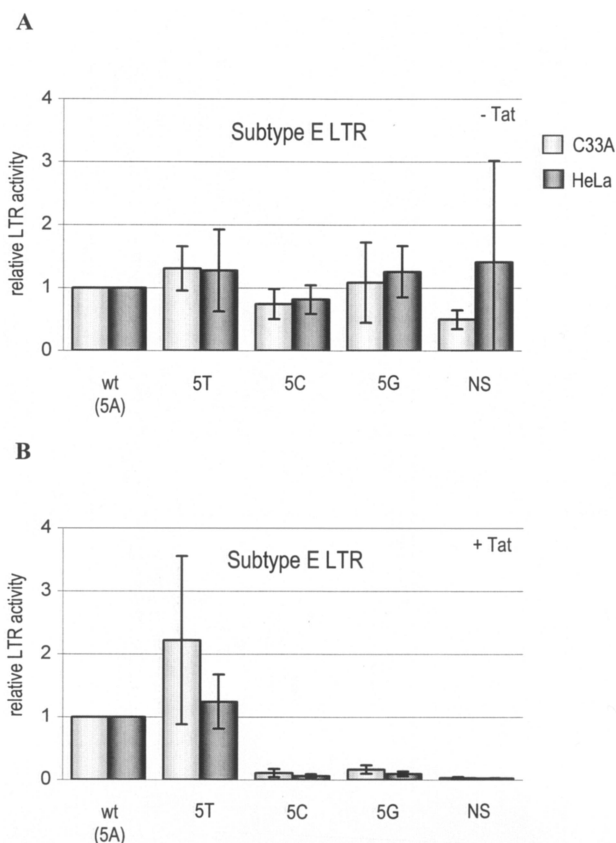


FIG. 3. Transcriptional activities of subtype E TATA box mutants. Shown are the relative luciferase activities of four TATA box mutant LTRs and the wt subtype E LTR. The activity of the wt TATA box sequence was arbitrarily set at 1 in assays without (A) and with (B) the transactivator protein Tat. Transfections were performed eight times in C33A and HeLa cells. Error bars indicate standard deviations.

Because only 1% of cellular pol II promoters have an A at position 3 (Table 2), the results seem to be incompatible with the original TATA box designation, in which the T/A polymorphism in subtypes B and E occurs at position 3. In addition, it has been shown that introduction of an A at position 3 in a model system leads to a complete loss of promoter activity (Table 3). This apparent contradiction can be solved by a 2-nucleotide upstream realignment of the HIV-1 TATA box to a CATA box. Consequently, the T/A polymorphism is present at position 5. The same sequence variation is present in cellular

TABLE 2. TATA box sequence variation in natural pol II cellular promoters

Nucleotide	Occurrence (%) ^a at TATA box position:						
	T (+1)	A (+2)	T (+3)	A (+4)	A (+5)	A (+6)	A (+7)
A	4	90	1	91	69	93	57
C	12	0	4	0	0	1	1
G	5	1	1	1	0	5	11
T	79	9	94	8	31	1	31

^a Occurrence of the four nucleotides in the TATA box at positions 1 to 7, based on 389 TATA box elements, mainly from vertebrate promoters obtained from TRANSFAC 6.0 (29). Boldface indicates the most favored residues.

TABLE 3. Promoter activities of all possible TATA box point mutants

wt nucleotide	TATA box activity ^a with the following nucleotide:						
	T (+1)	A (+2)	T (+3)	A (+4)	A (+5)	A (+6)	A (+7)
A	0.30	1.00	0	1.00	1.00	1.00	1.00
C	0.26	0.07	0.16	0.02	0.05	0.08	0.22
G	0.08	0.03	0.02	0	0.05	0.54	0.42
T	1.00	0.54	1.00	0.19	0.72	0.21	0.99

^a Activity relative to that of the wt (from reference 30). Boldface indicates the wild-type nucleotide (activity set at 1.00).

promoters with a prevalence of 31% T and 69% A (Table 2), and both variants are transcriptionally active (Table 3). Furthermore, G or C is never present at position 5 in pol II promoters (Table 2), and both are inactive in mutant cellular promoters (Table 3) and the HIV-1 LTR (Fig. 2 and 3). Thus, the 2-nucleotide realignment can explain the T/A polymorphism in subtypes B and E.

The T/A polymorphism at position 5 in the TATA box is compatible with virus replication. The first set of TATA mutations were inserted into the LAI molecular clone, a CXCR4-using primary isolate belonging to subtype B, and tested for virus replication on the SupT1 T-cell line (Fig. 4). Two mutants (5T and NS) were also inserted into the hybrid LAI molecular clone with the subtype E LTR. The results are fully consistent with the transient luciferase assays. The T/A sequence variation at position 5 of the TATA box is compatible with efficient virus replication, in the LTR contexts of both subtypes B and E. In contrast, the C/G virus variants are fully replication impaired, and the same result was obtained with the TATA box-inactivating mutant NS.

HIV-1 uses a CATA box for transcription and viral replication. A second set of mutations was made to confirm the hypothesis of a 2-nucleotide upstream realignment to the CATA box. These “new” position 1 and 2 nucleotides were mutated in the LTRs of subtypes B and E to determine their contribution to transcriptional activity. We show only the results of the LTR-luciferase assays with Tat. At position 1, all four nucleotides are compatible with efficient LTR transcription (Fig. 5). Position 2 is more important, as the 2G mutant reduced LTR activity to 41 and 46% in the subtype B and E contexts, respectively. The double mutant 1T2G further decreased LTR activity to 18% in subtype B, and transcriptional activity in the subtype E LTR was comparable to that of the 2G mutant. We also tested the separate 1T and 2G mutations in subtypes B and E for the effect on viral replication. The replication results are consistent with the transcription results; the 1T mutation does not affect virus replication, but the 2G mutation abolished viral replication (Fig. 6). The first position is indeed one of the most flexible sites in cellular TATA boxes, in terms of both the presence of all four nucleotides and the effect of mutations on transcriptional activity (Tables 2 and 3) (7, 30). The results presented for the HIV-1 LTR follow this pattern. The 2G mutation strongly affects LTR activity and viral replication, which is expected given the 1% occurrence in cellular pol II promoters and the negligible 3% activity of such a mutant in other studies (Tables 2 and 3).

We also mutated position 4 (the “old” position 2) in the

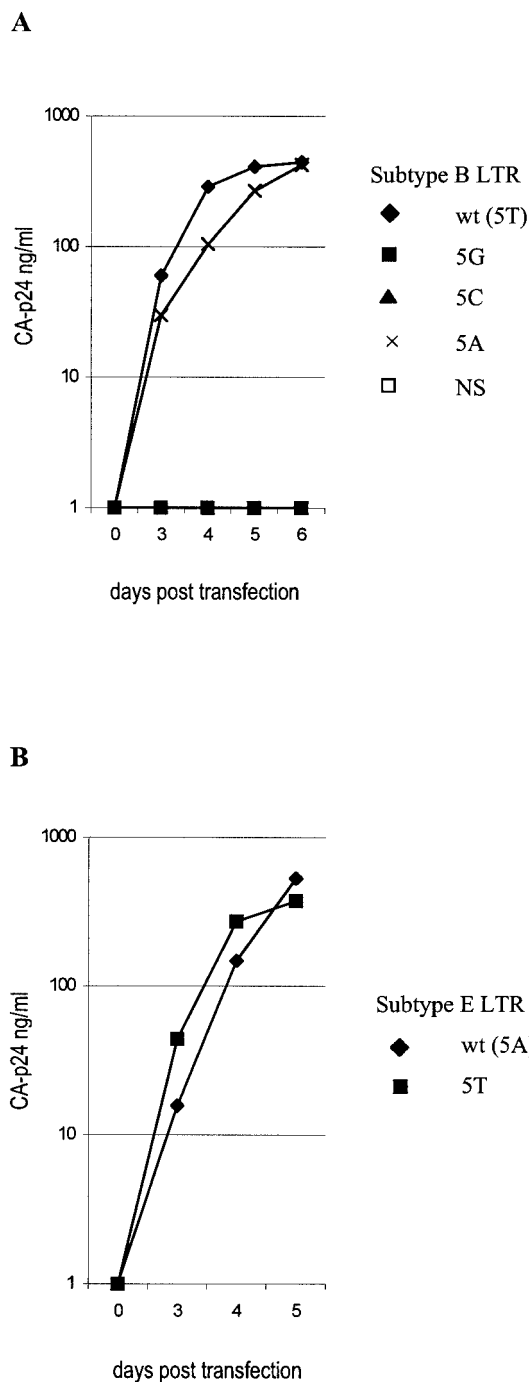


FIG. 4. Replication of TATA box mutant viruses in SupT1 cells. (A) Replication of the wt subtype B virus (LAI) and four TATA box mutants. (B) Replication of the hybrid LAI virus with the subtype E LTR and the 5T mutant.

CATA box from A to G, which reduced LTR activity to 12 and 19% for subtypes B and E, respectively (Fig. 5). This mutation abolished virus replication in both LTR contexts (Fig. 6). This result is expected for a position 4 nucleotide (Tables 2 and 3), although it does not provide evidence against the old position 2 classification.

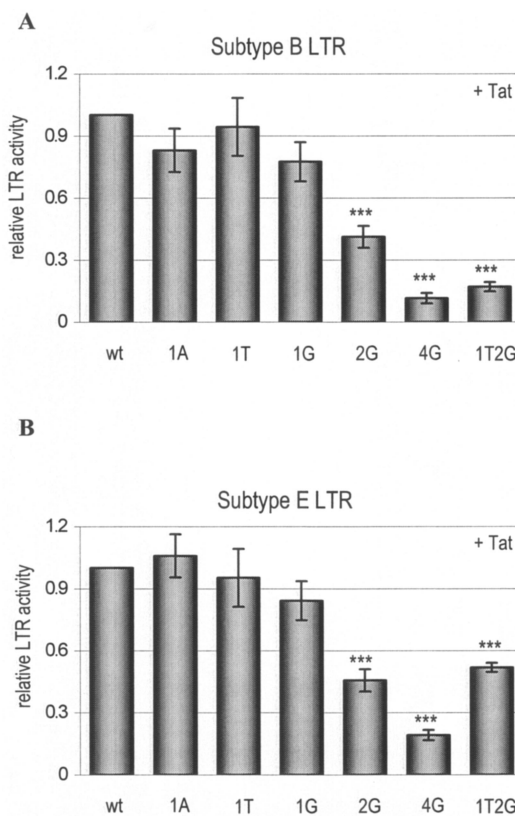


FIG. 5. Transcription is impaired through mutation of position 2 in the CATA box. Shown are six CATA box mutants in the subtype B and E contexts. C33A cells were transfected with the indicated LTR-luciferase plasmids and Tat plasmid. Luciferase activity was measured at 2 days posttransfection. All values were related to the wt activity, which was arbitrarily set at 1 in the subtype B (A) and subtype E (B) contexts. The significance of changes in promoter activity with respect to the “parental” CATA box was determined by using a paired *t* test (***, *P* < 0.001). Error bars indicate standard errors.

TATA box inactivation by mutation of position 2 can be rescued by a mutant TBP. To further validate the proposed 2-nucleotide upstream realignment to the CATA box, we tried to rescue LTR-luciferase activity of the 2G mutants by cotransfection of a corresponding mutant TBP (TBP_{AS}). TBP_{AS} contains three amino acid substitutions compared to TBP_{wt} and binds efficiently to the variant TG \overline{T} A \overline{T} AAA sequence (25). TBP_{AS} should thus be able to rescue transcription of the 2G mutants if the hypothesis of the realigned CATA box is correct. We were able to significantly increase transcription of both the 2G and 1T2G mutants in the subtype B and E LTRs by cotransfection of TBP_{AS} (Fig. 7). The activity of the 4G mutation in the subtype E CATA box was slightly improved by TBP_{AS}, although activity remained very low. The LTR activity of the 4G mutation in subtype B was not significantly increased by TBP_{AS}. The combined results indicate that TBP indeed interacts with the realigned CATA box and that the 2-nucleotide upstream shift is justified.

The HIV-1 CATA box can be improved for viral replication by a point mutation. We have demonstrated that the subtype E LTR promoter is functional despite the T/A sequence polymorphism because it uses the CATAAAA box. It is neverthe-

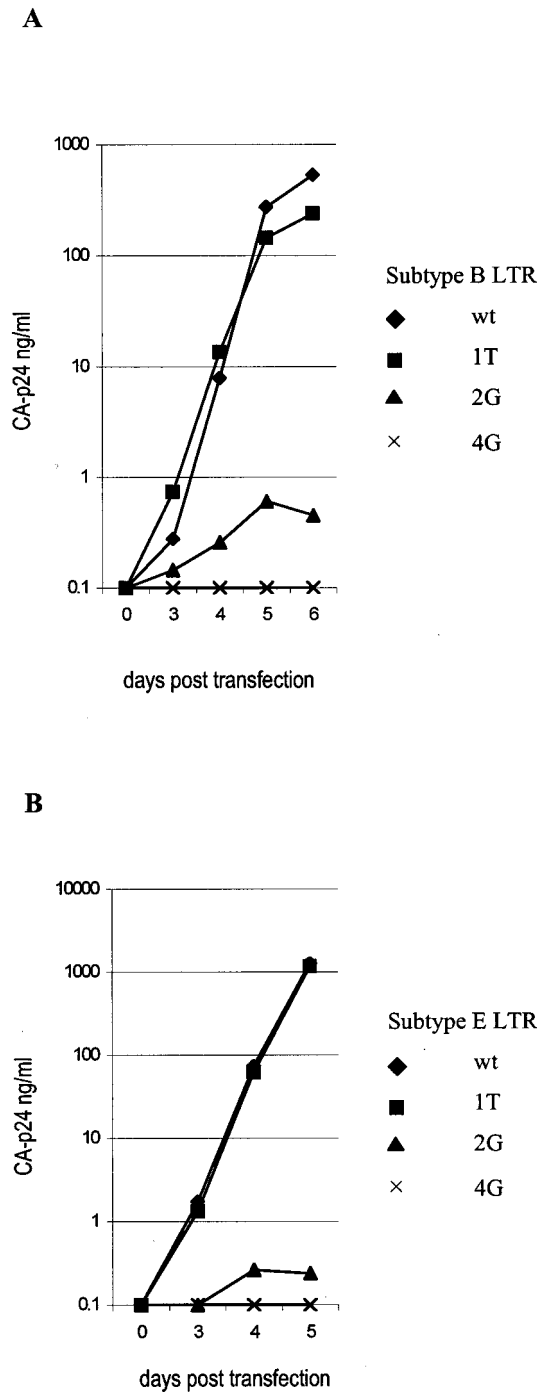


FIG. 6. Replication is impaired through mutation of position 2 in the CATA box. (A) Replication of the subtype B virus with a wt TATA box and three subtype B TATA box mutants. (B) Replication of the subtype E LTR virus with a wt TATA box and three subtype E TATA box mutants.

less striking that this T-to-A change at position 5 is seen exclusively in this subtype and some recombinant AG strains, which suggests that it is not a selectively neutral mutation. Replication assays are often not sensitive enough to determine small fitness differences between two virus variants. We therefore performed pairwise competition experiments with four

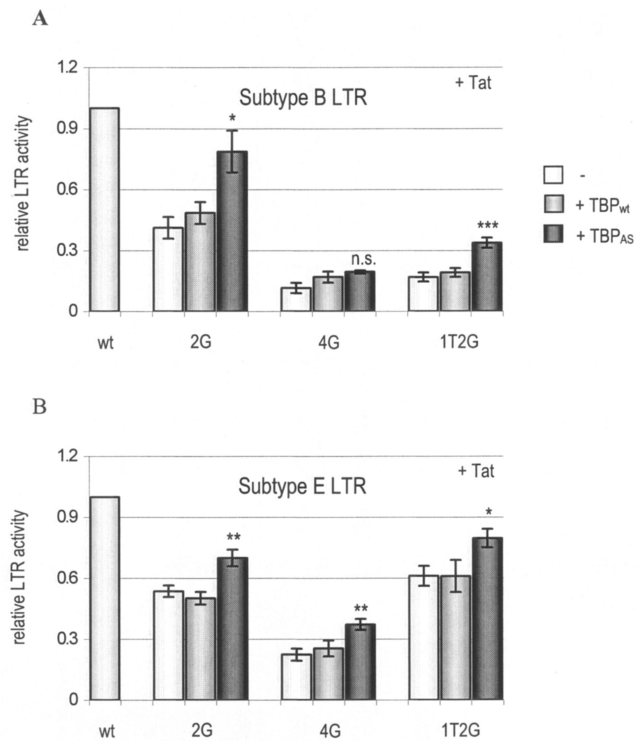


FIG. 7. The mutant TBP_{AS} rescues transcription of position 2 CATA box mutants. Shown are three CATA box mutants in the subtype B (A) and E (B) contexts. Control transfections (-) and cotransfections with TBP_{wt} and TBP_{AS} were performed. No significant differences were observed between control transfections and cotransfection with TBP_{wt}. There were significant differences only between cotransfections with TBP_{AS} and TBP_{wt}, which were determined by a paired *t* test (*, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001; n.s., not significant). Error bars indicate standard errors.

active virus mutants against the wt “parental” virus. The 5A mutation in subtype B changes the CATATAA box (subtype B wt) into the CATAAAA box (subtype E wt), but this mutant is out-competed by the parental virus and is thus less fit (Table 4). The same comparison in the subtype E context yields the identical result; the CATATAA variant out-competes the CA TAAAA virus (Table 4). The latter result is surprising because it means that the 5T mutant is more fit than the parental E virus. The CATA box of subtype B can also be improved for its replication function. The 1T mutant changes the CATATAA box of subtype B into the sequence TATATAA, which appar-

TABLE 4. Relative fitnesses of four viruses with respect to their wt “parental” virus (subtype B or E)

	wt	Mutant	Relative fitness ^a (mean ± SE)
Subtype B (CATATAA)		5A (CATAAAA)	0.97 ± 0.008
		1T (TATATAA)	1.04 ± 0.025
Subtype E (CATAAAA)		5T (CATATAA)	1.03 ± 0.007
		1T (TATAAAA)	0.96 ± 0.006

^a Fitness of the mutant measured against the corresponding wt virus (subtype B or E) containing a subtype-specific LTR.

ently increases viral fitness (Table 4). However, this effect is context dependent, because the 1T mutation in the subtype E CATA box (CATAAAA to TATAAAA) decreases viral fitness. These combined results indicate that it is possible to increase the virus replication rate by distinct point mutations in the CATA box (subtype E, CATAAAA [wt] < CATATAA; subtype B, CATATAA [wt] < TATATAA). Therefore, since there is room for improvement, the CATA box of HIV-1 seems to be tuned for a suboptimal replication rate, which might, however, be optimal *in vivo*.

DISCUSSION

There are apparent differences in the genetic make-ups of the LTR transcriptional promoters of the HIV-1 subtypes that are observed in the current pandemic. An intriguing sequence polymorphism is present in the TATA boxes of all subtype E isolates and several recombinant AG strains, which contain the sequence TAAAAGC, whereas all other subtypes have the regular TATAAGC box. The subtype E TAAAAGC box is remarkable because a T-to-A change at the third position is known to abolish all transcriptional activity in other promoters (7, 30). In this study, we provide an unexpected explanation for this phenomenon. We demonstrate that the TATA boxes of all HIV-1 viruses have been mapped incorrectly. The true motif starts 2 nucleotides further upstream (CATAAAA), such that the subtype E polymorphism is at position 5 and therefore is compatible with promoter activity. The evidence for the realigned CATA box is threefold. First, we determined that the T/A polymorphism does not affect LTR promoter activity in subtypes B and E but that a C or G substitution abolishes promoter activity and virus replication. This pattern is not consistent with the "old" position 3 but is fully consistent with the "new" position 5 in the CATATAA box. Second, the new position 2 nucleotide in the CATATAA box significantly contributes to the transcriptional activity. The new C at position 1 can be replaced without a significant loss of function, which is somewhat unexpected for nucleotide G at position 1 (Tables 2 and 3). The stricter sequence specificity of the A at position 2 is fully consistent with the requirements of this TATA box residue. Third, we were able to partially rescue the activity of the 2G promoter mutants by cotransfection of a mutant TBP that selectively binds to the variant TGTAAAA box. These combined results demonstrate that the true TATA box of HIV-1 is in fact a CATA box that starts 30 bp upstream of the transcription start site. This also means that alternative explanations, such as compensatory changes elsewhere in LTR-TAR or a TATA-less promoter, can be dismissed (9, 13).

A nonconsensus TATA box is present in many cellular pol II promoters (Table 2), and their prevalence may even be underestimated due to "misalignment" as we have shown here. Within the group of retroviruses there are several species that also have a nonconsensus TATA box. Several types of simian immunodeficiency virus (e.g., SIVcpz) possess the same CAT AAAA box as subtype E of HIV-1. Murine endogenous retrovirus and porcine endogenous retrovirus have an A at position 1 (AATAAAA), and bovine leukemia virus has the motif GATAAAT. HIV-2 and human T-cell leukemia virus type 1 contain the consensus TATA box sequence TATAAAA. This retrovirus survey confirms that some variation in the genetic

make-up of the TATA box is possible. Besides the T/A polymorphism in the CATA box, there are many other differences in the LTR transcriptional promoters of the HIV-1 subtypes. For example, subtype C contains three NF- κ B sites whereas most other subtypes possess two NF- κ B sites (14), and subtype E contains a unique 1-nucleotide deletion in the upstream NF- κ B site that converts this motif into a GA binding protein site (28). We have recently shown that the TFBS within the U3 region determine the virus replication rate in a subtype-specific manner (26). In addition, we showed that the activation status of the host cell modulates virus replication in a subtype-specific manner. The origin of these fitness differences of the HIV-1 subtypes is unclear, although adaptation to different cellular environments may be a likely explanation. In this study, we determined a small but significant fitness advantage for viruses with CATATAA over CATAAAA in sensitive competition experiments. This result was obtained for the LTR contexts of both subtypes B and E. The subtype E wt CATAAAA box is thus less fit for virus replication than the standard CATATAA box sequence of all other subtypes. Interestingly, we also determined that fitness of the subtype B virus can be increased by the 1T mutation (CATATAA to TATATAA). Thus, *in vitro* replication of both subtype viruses can be improved by manipulation of the CATA box. Similar alterations at positions 1 and 5 in the CATA box were previously identified as escape mutations in studies with transcriptionally impaired HIV-1 variants lacking either all Sp1 motifs or all Sp1 motifs and one NF- κ B site (23a).

If point mutations in the HIV-1 CATA box can indeed improve viral fitness, why then have these mutations not been fixed in their respective subtype populations? A possibility is that the replication rate of HIV-1 is selected for optimal performance, which may coincide with a lower-than-maximal replication rate. For instance, transcription levels in virus-infected cells may be fine-tuned to prevent overexpression of potentially toxic proteins. Consistent with this idea is the observation that a modified HIV-1 with a very strong artificial promoter evolves over time, in cell culture, to a variant with a moderately active promoter, coinciding with increased virus replication (12). There may be many additional reasons for fine-tuning of LTR activity, of which we will mention three examples. First, a too-high replication rate may disturb the complex sequence of events during HIV-1 protein production and virion assembly that eventually results in the budding of infectious viral particles from the cell. Second, the level of TBP expression may vary between different cell types. The CATA box motif of subtypes B and E may therefore be differentially optimized for transcription in a particular cell type, e.g., macrophages or monocytes. Third, a suboptimal CATA box may be more sensitive to transcriptional silencing and may thus have relevance for the issue of latently infected cells, e.g., nonactivated T cells.

ACKNOWLEDGMENTS

We thank Marc Timmers and Lloyd Pereira for supplying us with the wild-type and mutant TBP expression plasmids. We thank Maarten Boerlijst and Fedde Groot for discussions and comments on the manuscript.

This research was supported by NWO-ALW (project 811.35.001).

REFERENCES

1. Albright, S. R., and R. Tjian. 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**:1–13.
2. Auersperg, N. 1964. Long-term cultivation of hypodiploid human tumor cells. *J. Natl. Cancer Inst.* **32**:135–163.
3. Berkhout, B., A. Gatignol, A. B. Rabson, and K. T. Jeang. 1990. TAR-independent activation of the HIV-1 LTR: evidence that tat requires specific regions of the promoter. *Cell* **62**:757–767.
4. Berkhout, B., and K. T. Jeang. 1992. Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **66**:139–149.
5. Blackard, J. T., B. Renjifo, W. Fawzi, E. Hertzmark, G. Msamanga, D. Mwakagile, D. Hunter, D. Spiegelman, N. Sharghi, C. Kagoma, and M. Essex. 2001. HIV-1 LTR subtype and perinatal transmission. *Virology* **287**:261–265.
6. Das, A. T., B. Klaver, and B. Berkhout. 1999. A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J. Virol.* **73**:81–91.
7. Hoopes, B. C., J. F. LeBlanc, and D. K. Hawley. 1998. Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J. Mol. Biol.* **277**:1015–1031.
8. Hu, D. J., S. Vanichseni, T. D. Mastro, S. Raktham, N. L. Young, P. A. Mock, S. Subbarao, B. S. Parekh, L. Srisuwanvilai, R. Sutthent, C. Wasi, W. Heneine, and K. Choopanya. 2001. Viral load differences in early infection with two HIV-1 subtypes. *AIDS* **15**:683–691.
9. Jeeninga, R. E., M. Hoogenkamp, M. Armand-Ugon, M. de Baar, K. Verhoef, and B. Berkhout. 2000. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J. Virol.* **74**:3740–3751.
10. Kaleebu, P., A. Ross, D. Morgan, D. Yirrell, J. Oram, A. Rutebemberwa, F. Lyagoba, L. Hamilton, B. Biryahwaho, and J. Whitworth. 2001. Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. *AIDS* **15**:293–299.
11. Kanki, P. J., D. J. Hamel, J. L. Sankale, C. Hsieh, I. Thior, F. Barin, S. A. Woodcock, A. Gueye-Ndiaye, E. Zhang, M. Montano, T. Siby, R. Marlink, I. NDooye, M. E. Essex, and S. Mboup. 1999. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J. Infect. Dis.* **179**:68–73.
12. Marzio, G., M. Vink, K. Verhoef, A. de Ronde, and B. Berkhout. 2002. Efficient human immunodeficiency virus replication requires a fine-tuned level of transcription. *J. Virol.* **76**:3084–3088.
13. Montano, M. A., C. P. Nixon, and M. Essex. 1998. Dysregulation through the NF- κ B enhancer and TATA box of the human immunodeficiency virus type 1 subtype E promoter. *J. Virol.* **72**:8446–8452.
14. Montano, M. A., V. A. Novitsky, J. T. Blackard, N. L. Cho, D. A. Katzenstein, and M. Essex. 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J. Virol.* **71**:8657–8665.
15. Naghavi, M. H., M. C. Estable, S. Schwartz, R. G. Roeder, and A. Vahlne. 2001. Upstream stimulating factor affects human immunodeficiency virus type 1 (HIV-1) long terminal repeat-directed transcription in a cell-specific manner, independently of the HIV-1 subtype and the core-negative regulatory element. *J. Gen. Virol.* **82**:547–559.
16. Naghavi, M. H., S. Schwartz, A. Sonnerborg, and A. Vahlne. 1999. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. *AIDS Res. Hum. Retroviruses* **15**:1293–1303.
17. Ohler, U., and H. Niemann. 2001. Identification and analysis of eukaryotic promoters: recent computational approaches. *Trends Genet.* **17**:56–60.
18. Patikoglou, G. A., J. L. Kim, L. Sun, S. H. Yang, T. Kodadek, and S. K. Burley. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**:3217–3230.
19. Peden, K., M. Emerman, and L. Montagnier. 1991. Changes in growth properties on passage in tissue culture of viruses derived from infectious molecular clones of HIV-1LAI, HIV-1MAL, and HIV-1ELI. *Virology* **185**:661–672.
20. Pereira, L. A., K. Bentley, A. Peeters, M. J. Churchill, and N. J. Deacon. 2000. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res.* **28**:663–668.
21. Pugh, B. F. 2000. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**:1–14.
22. Renjifo, B., W. Fawzi, D. Mwakagile, D. Hunter, G. Msamanga, D. Spiegelman, M. Garland, C. Kagoma, A. Kim, B. Chaplin, E. Hertzmark, and M. Essex. 2001. Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. *J. Hum. Virol.* **4**:16–25.
23. Rinke de Wit, T. F., A. Tsegaye, D. Wolday, B. Hailu, M. Aklilu, E. Sanders, M. Hagos, A. Kliphuis, G. Pollakis, A. Krol, R. Geskus, F. Miedema, J. Goudsmit, R. Coutinho, and A. L. Fontanet. 2002. Primary HIV-1 subtype C infection in Ethiopia. *J. Acquir. Immune Defic. Syndr.* **30**:463–470.
- 23a. Ross, E. K., A. J. Buckler-White, A. B. Rabson, G. Englund, and M. A. Martin. 1991. Contribution of NF- κ B and Sp1 binding motifs to the replicative capacity of human immunodeficiency virus type 1: distinct patterns of viral growth are determined by T-cell types. *J. Virol.* **65**:4350–4358.
24. Smith, C. D., M. Shatsky, P. S. Cohen, R. Warnke, M. P. Link, and B. E. Glader. 1984. Monoclonal antibody and enzymatic profiles of human malignant T-lymphoid cells and derived cell lines. *Cancer Res.* **44**:5657–5662.
25. Strubin, M., and K. Struhl. 1992. Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell* **68**:721–730.
26. van Opijnen, T., R. E. Jeeninga, M. C. Boerlijst, G. P. Pollakis, V. Zetterberg, M. Salminen, and B. Berkhout. 2004. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. *J. Virol.* **78**:3675–3683.
27. Verhoef, K., M. Koper, and B. Berkhout. 1997. Determination of the minimal amount of Tat activity required for human immunodeficiency virus type 1 replication. *Virology* **237**:228–236.
28. Verhoef, K., R. W. Sanders, V. Fontaine, S. Kitajima, and B. Berkhout. 1999. Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF- κ B enhancer element into a GABP binding site. *J. Virol.* **73**:1331–1340.
29. Wingender, E., X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhauser, M. Pruss, F. Schacherer, S. Thiele, and S. Urbach. 2001. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29**:281–283.
30. Wobbe, C. R., and K. Struhl. 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Mol. Cell. Biol.* **10**:3859–3867.