Supporting Information

A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter

Rosalba Perrone[a], Matteo Nadai[a], Ilaria Frasson[a], Jerrod A. Poe[b], Elena Butovskaya[a], Thomas E. Smithgall[b], Manlio Palumbo[c], Giorgio Palù[a], Sara N. Richter[a*]

[a]Department of Molecular Medicine, via Gabelli 63, 35121 Padua, Italy; [b]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [c]Department of Pharmaceutical and Pharmacological Sciences, via Marzolo 5, 35131 Padua, Italy.

Supporting Figure 1

See landscape format file

Figure S1. Sequence and base conservation of the G-rich HIV-1 LTR region spanning -105/-48 nts. A) G bases that may be involved in G-quadruplexes are shown in bold. G-tracts are consecutively numbered (1-6 for GGG- or GGGG-tracts and 3', 3", 6' for GG-tracts). The orange and magenta lines above the sequence indicate the bases involved in NF-kB and Sp1, respectively, transcription factor binding. The sequences that may form G-quadruplexes with two stacked G-quartets are indicated by Arabic numbers and differently colored brackets. B) Base conservation among 953 HIV-1 strains (24, 485, 119 and 325 strains of subtypes A, B, C and others, respectively). G-bases are shown in red, non-G-bases are shown in lilac. Percentages of conservation of G-bases are shown above corresponding bars.

Supporting Figure S2

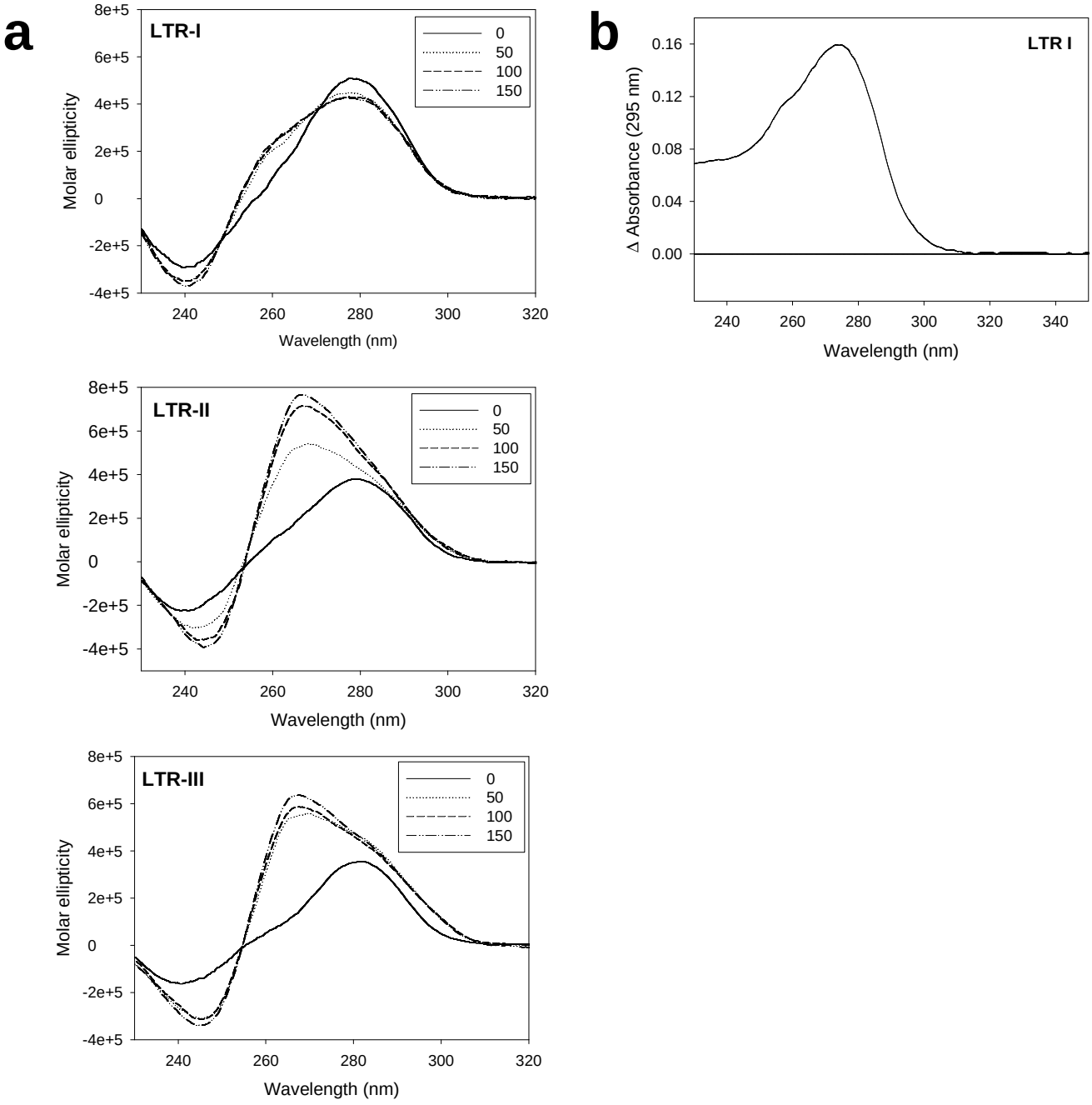


Figure S2. Spectroscopic data of LTR I and LTR III putative G-quadruplex forming sequences. A) CD spectra of each oligonucleotide (4 μM) in the presence of increasing concentration of K^+ (0-150 mM). B) TDS at 295 nm of LTR I.

Supporting Figure S3

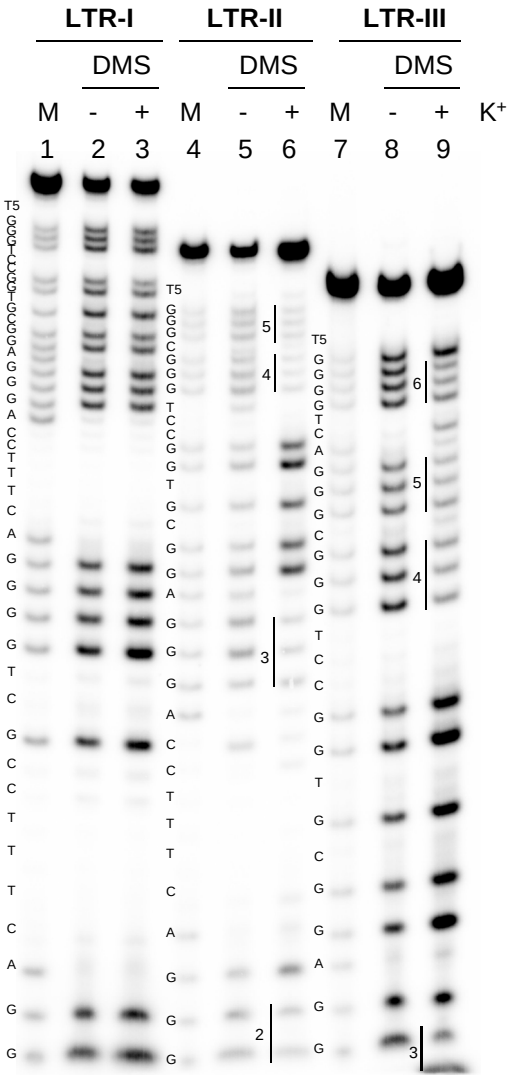


Figure S3. DMS-protection assay of LTR-I, LTR-II and LTR-III. LTR sequences were heat denatured, cooled down in the presence or absence of K^+, and treated with DMS before cleavage induction at the G alkylated sites with hot piperidine. Base sequences are provided on the left of each oligonucleotide. Protected G-tracts are indicated with vertical lines and corresponding numbers (see Fig. 1). The ¤ symbol indicate overexposed bases and the * symbol indicate non-protected bases within a protected G-tract. M indicate marker lanes obtained with the Maxam&Gilbert sequencing protocol.

Supplementary Figure S4

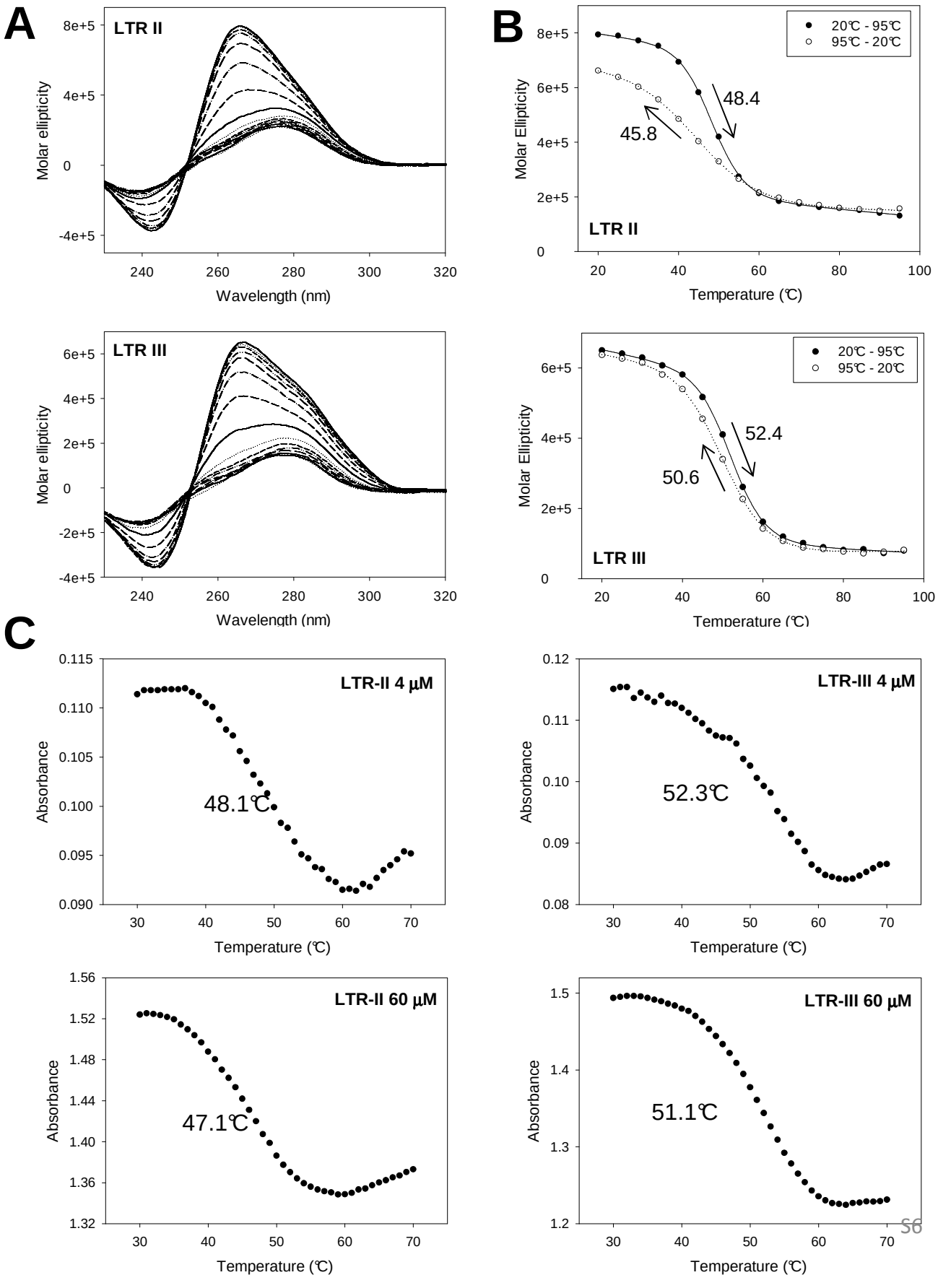


Figure S4. Thermal stability of LTR-II and LTR-III. A) LTR-II and LTR-III oligonucleotides (3-6 μM) were treated at increasing/decreasing temperature (20-95°C 2°C/min) and CD spectra recorded. B) Molar ellipticity values recorded at 265 nm were plotted against temperature and fitted with the van't Hoff equation to extrapolate T_m values. Down- and up-pointing arrows indicate denaturation and renaturation experiments, respectively. T_m values are indicated aside the corresponding melting curve. C) LTR-II and LTR-III oligonucleotides at low (4 μM) and high concentration (60 μM) were treated at increasing temperature (20-95°C) and UV absorbance spectra recorded. Absorbance values recorded at 295 nm were plotted against temperature to extrapolate T_m values, which are indicated aside the corresponding melting curve.

Supporting Figure S5

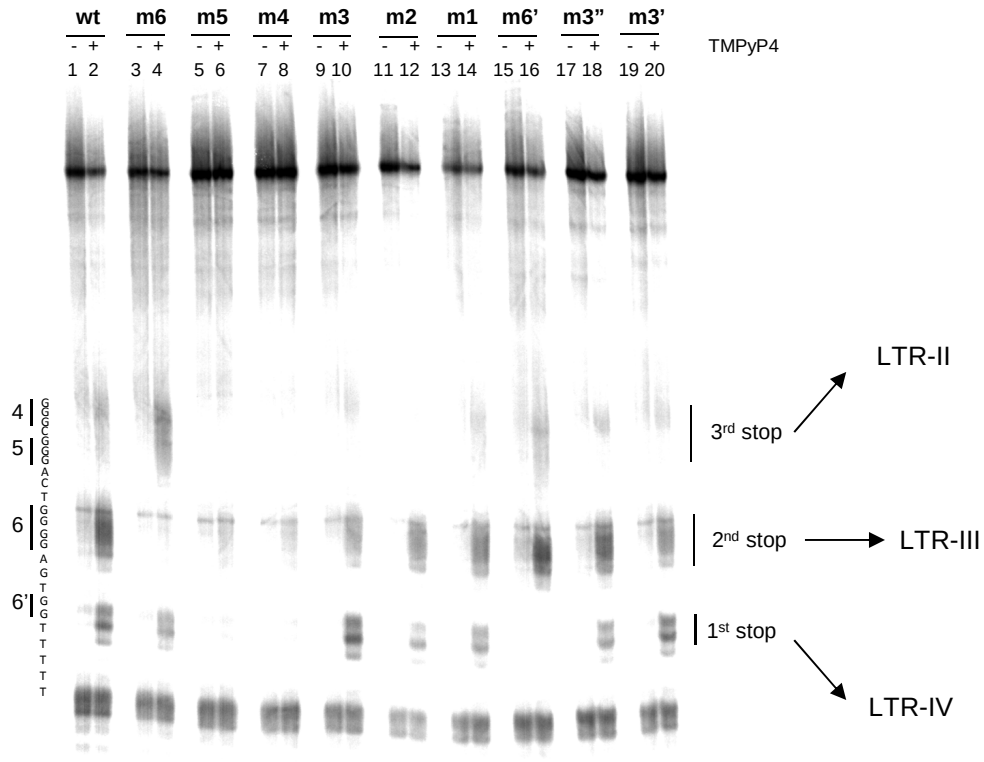


Figure S5. Taq polymerase stop assay on wild-type and mutant FL LTR templates in the presence of 100 mM K^+ and 100 nM TMPyP4.

Supporting Figure S6

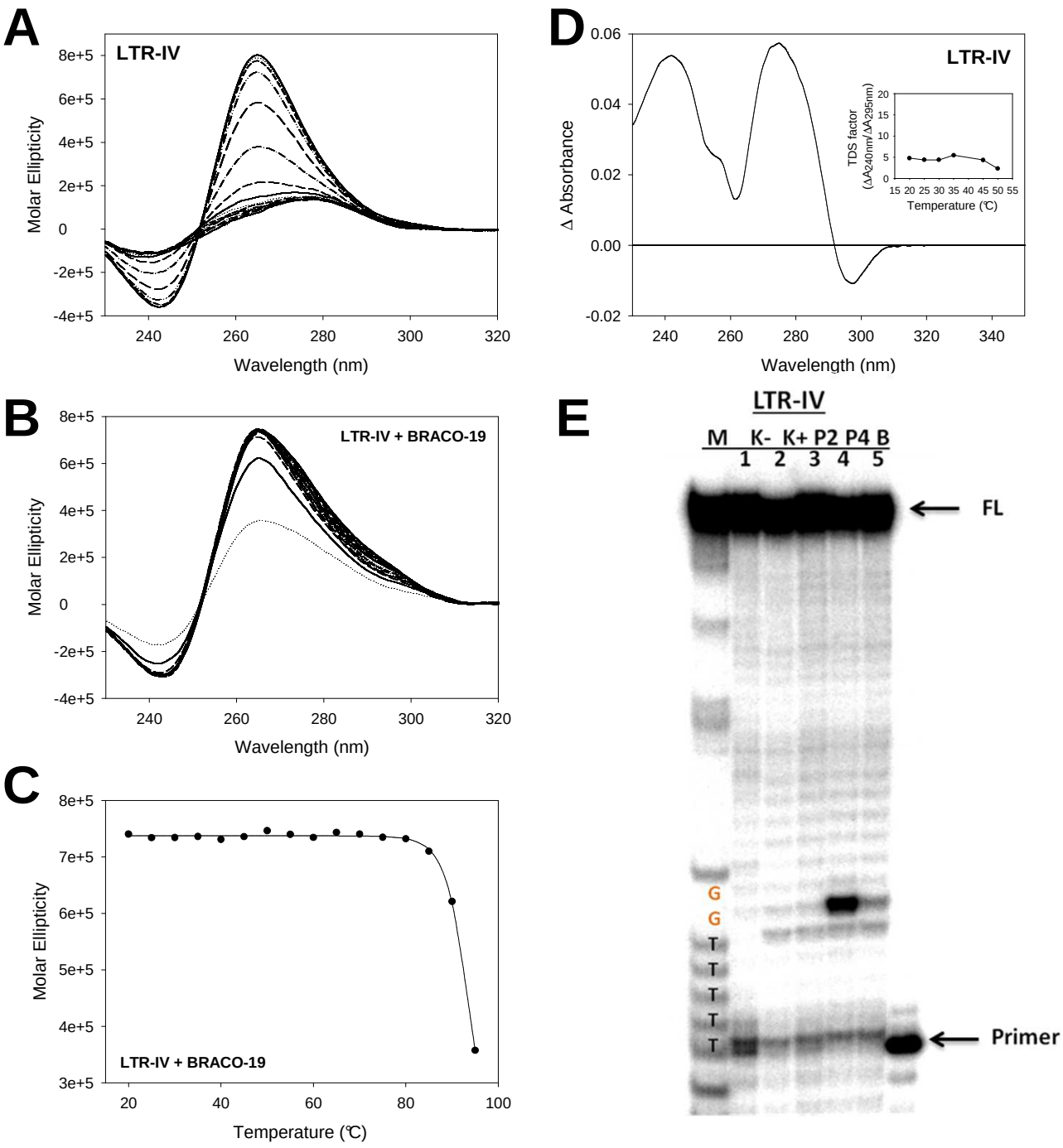


Figure S6. Thermal stability of LTR-IV in A) the absence and B) presence of BRACO-19. LTR-IV oligonucleotide was treated at increasing temperature (20-95°C) and CD spectra recorded. C) Molar ellipticity values recorded at 265 nm were plotted against temperature and fitted with the van't Hoff equation to extrapolate T_m values. D) Thermal differential spectra (TDS) at 295 nm of LTR-IV. TDS factor graph is shown in the inset. E) Taq polymerase stop assay performed at 47°C in the absence (K-), presence (K+) of K^+, and in the presence of TMPyP2 (P2), TMPyP4 (P4) or BRACO-19 (B). FL indicates the full-length polymerase product. The 3'-end sequence is shown on the left.

Supplementary Figure S7

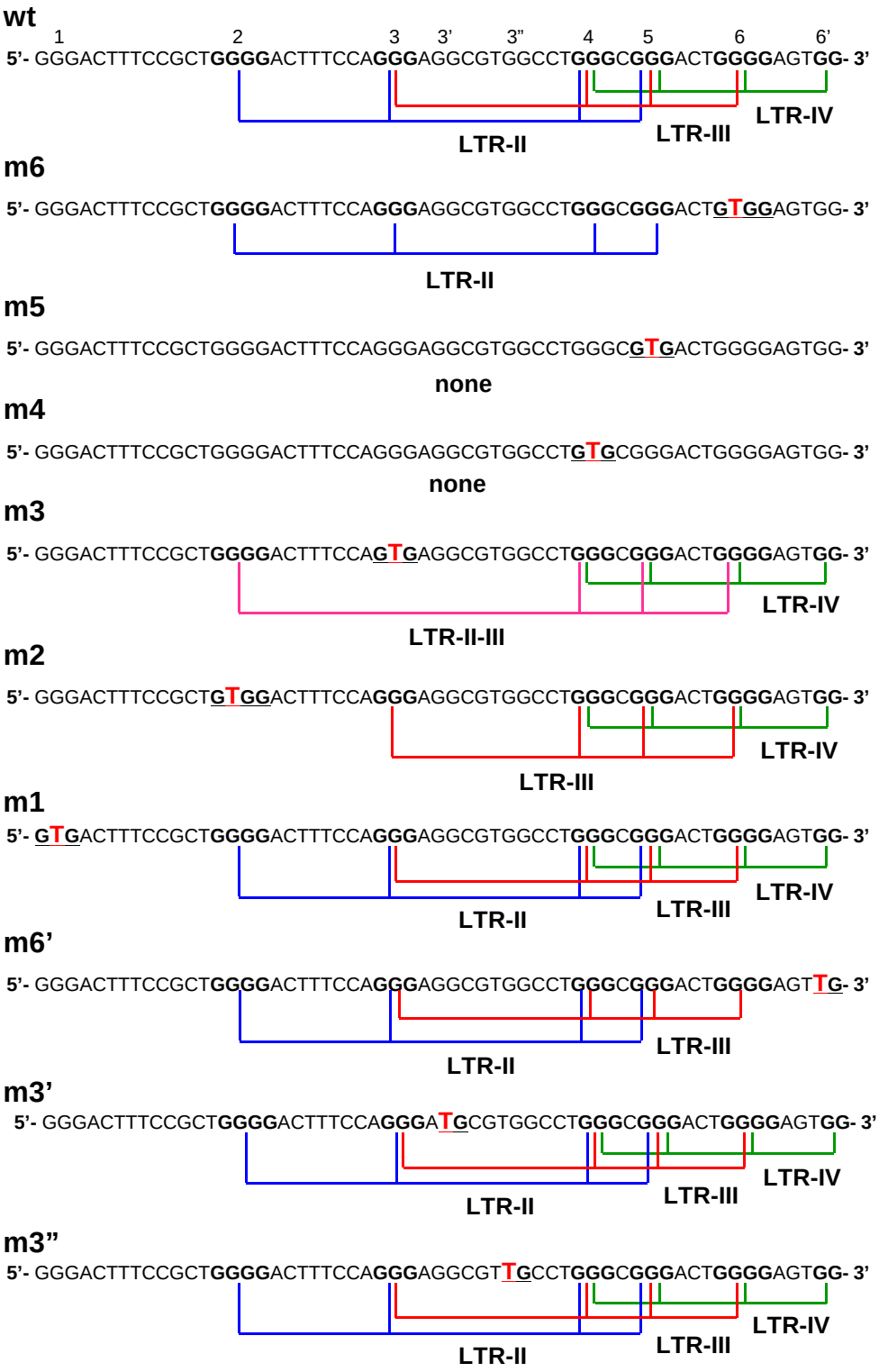
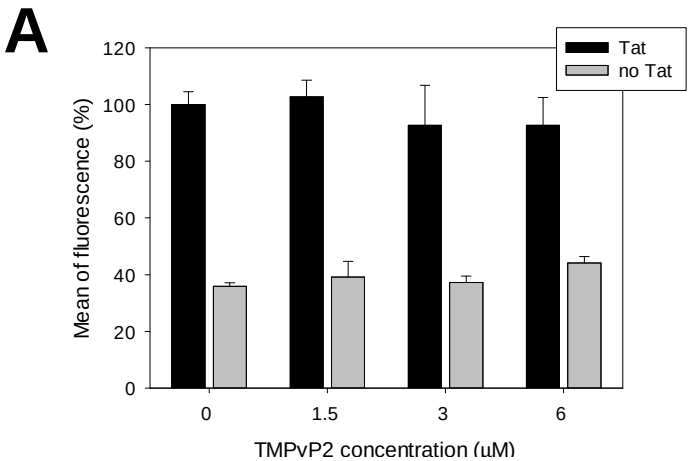
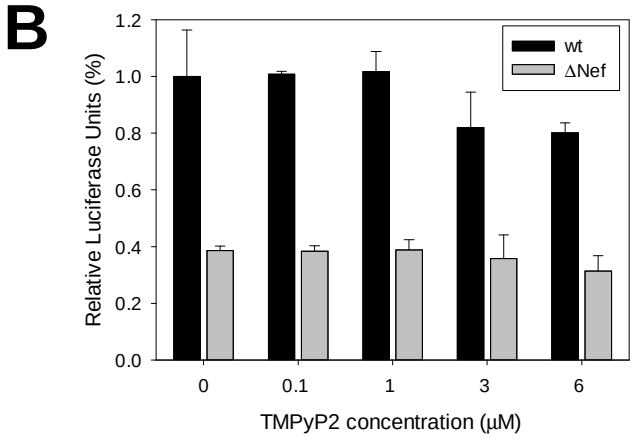


Figure S7. G-quadruplexes forming in each of the examined mutant FL-LTRs. Brackets indicate G-tracts involved in the relevant G-quadruplex. Blue, red and green brackets indicate LTR-II, LTR-III, LTR-IV G-quadruplexes, respectively. Mutated bases are shown in red. LTR-II-III is an alternative structure that forms involving G-tracts 2, 4, 5 and 6 when G-tract 3 is unavailable.

Supporting Figure S8





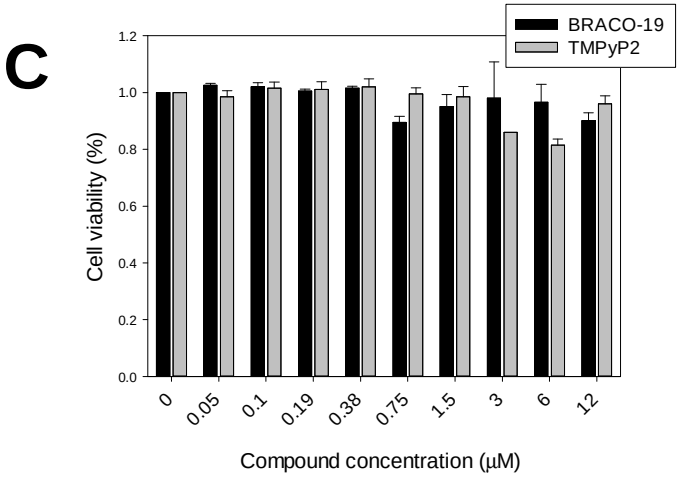


Figure S8. Effect of TMPyP2 on the LTR promoter and in infected cells. A) GFP mean of fluorescence of cells transfected with the wt LTR-GFP plasmid and treated with increasing concentration (1.5-6.0 μM) of TMPyP2 in the absence or presence of Tat. In all data sets: $n = 3$, mean ± s.d., Student's *t*-test. No significant difference was found. B) TZM-bl cells were infected with wild-type (black bars) and ΔNef (grey bars) HIV NL4-3 in the presence of the negative control G-4 ligand, TMPyP2. After 48 h, levels of gene expression was assessed as relative luciferase activity in infected cells. Results are shown relative to the wild-type control cells incubated with only the carrier solvent (DMSO) ± SEM (n = 3).C) Cytotoxicity of BRACO-19 and TMPyP2 on TZM-bl cells was assessed via the Cell-Titer Blue assay (Promega). TZM-bl cells were incubated with the indicated concentrations of compounds for 48 h and cell viability was assessed via the Cell-Titer Blue assay relative to control cells incubated with carrier solvent. Assays were done in triplicate.

Table S1. Properties of G-quadruplex forming sequences in the -105/-48 HIV-1 LTR U3 region.

Sequence name	# G-quartets	# linker nucleotides (min-max)			G-score
		loop I	loop II	loop III	
LTR-I	3	10	8	11	69-70
LTR-II	3	8-9	11	1	62-68
LTR-III	3	11	1	3-4	62
LTR-4	2	10-11	0	8-9	10-11
LTR-5	2	9-11	7-9	1-2	13
LTR-6	2	0	8-9	12	12-13
LTR-7	2	8-11	1-2	3	15
LTR-8	2	1-2	3	3-4	18-20
LTR-9	2	3	3-4	1-3	18-21
LTR-10	2	3	1-3	3-4	13-21
LTR-11	2	1-3	3-4	0	17-18
LTR-12	2	1-3	3-5	3-4	13-20

Table S2. Oligonucleotides used in this study. Mutated based in the Taq polymerase templates are shown underlined and in bold.

Application	Name	Sequence (5'→3')	Vector
CD, footprinting	FL-LTR	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGCGGGACTGGGGAGTGGTTTTT	na
	LTR-I	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGTTTTT	na
	LTR-II	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCG GGTTTTT	na
	LTR-III	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGTTT TT	na
	LTR-(II+III)	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCG GGACTGGGGTTTTT	na
Taq polymerase stop assay	LTR G4 Taq primer	GGCAAAAAGCAGCTGCTTATATGCAG	na
	LTR G4 LTR-I Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGTTTTTCTGCATATAAGCAGCTGCTTTT TGCC	na
	LTR G4 LTR-II Taq	TTTTTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCG GGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC	na
	LTR G4 LTR-III Taq	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGTTT TTCTGCATATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTGCA TATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL m1 Taq	TTTTTG**T**GACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCC**T**GGGCGGGACTGGGGAGTGGTTTTTCTGCA TATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL m2 Taq	TTTTTGGGACTTTCCGCTG**T**GGACTTTCCAGGGAGGC GTGGCCTGGGCGGGACTG**G**GGAGTGGTTTTTCTGCA TATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL m3 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAG**T**GAGGC GTGGCCTGGGCGGGACTGGGGAGTGGTTTTT**C**TGCA TATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL m3' Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGA**T**GC GTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTG**C**A TATAAGCAGCTGCTTTTTGCC	na
	LTR G4 FL m3" Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GT**T**GCCTGGGCGGGACTGTGGAGTGGTTTTTCTGCAT AT**A**AGCAGCTGCTTTTTGCC	na
	LTR G4 FL m4 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTG**T**GCGGGACTGGGGAGTGGTTTTTCTGCA TATAAGCA**G**CTGCTTTTTGCC	na
	LTR G4 FL m5 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGCG**T**GACTGGGGAGTGGTTTTTCTGCA TATAAGCAGCTG**C**TTTTTGCC	na
	LTR G4 FL m6 Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGCGGGACTG**T**GGAGTGGTTTTTCTGCA TATAAGCAGCTGCTTTTTG**C**C	na
	LTR G4 FL m6' Taq	TTTTTGGGACTTTCCGCTGGGGACTTTCCAGGGAGGC GTGGCCTGGGCGGGACTGGGGAGT**T**GTTTTTCTGCA TATAAGCAGCTGCTTTTTGCC	na
LTR cloning for reporter assays	LTR-Xho I	GGGCCCCTCGAGCCCTGATTGGCAGAAYTACACACC AGG	pGL4.10-Luc2/LTR-wt
	LTR-Hind III	GGGCCCAAGCTTCCTGCGTCGAGAGAGCTYCTCTGG	
LTR mutants cloning for reporter assays	pr m4a	CCAGGGAGGCGTGGCCTGTGCGGGACTGGGGAGTG GCG	pGL4.10-Luc2/LTR-m4
	pr m4b	CGCCACTCCCCAGTCCCGCACAGGCCACGCCTCCCT GG	