An Evolved RNA Recognition Motif That Suppresses HIV-1 Tat/TAR-Dependent Transcription

David W. Crawford,[1] Brett D. Blakeley,[1] Po-Han Chen,[3] Chringma Sherpa,[4] Stuart F.J. Le Grice,[4] Ite A. Laird-Offringa,[3] and Brian R. McNaughton*[,1,2]

[1]Department of Chemistry, Colorado State University, Fort Collins, Colorado, USA.

[2]Department of Biochemistry & Molecular Biology, Colorado State University, Fort Collins, Colorado, USA.

[3]Department of Surgery and Department of Biochemistry & Molecular Biology, USC/Norris Comprehensive Cancer Center, Keck School of Medicine, Los Angeles, California, USA.

[4]Basic Research Laboratory, National Cancer Institute, Frederick, Maryland, USA

*corresponding author: brian.mcnaughton@colostate.edu

Table of Contents

Materials and Instrumentation

Materials	S3
Instrumentation	S3

Additional Experimental Data

Protein library preparation	S4
β2- β3 Loop Library Construction	S4
Yeast Display Screening	S5
Table S1. Experimental conditions for yeast display	S5
Diversification of the 2nd generation library	S5
Protein Purification	S6
ELISA	S7
SPR	S8
Table S2. Kinetic data from SPR experiments	S8
ITC	S9
RNA preparation for SHAPE	S9
SHAPE and RNA migration experiments	S9
Transcription assay	S10
Figure S1. SHAPE RNA folding analysis	S11
Figure S2. Gel-shift data for TBP 6.6 or TBP 6.7 and HIV-1 5'-UTR	S12
Figure S3 and S4. SHAPE data using 4x TBP 6.6 or 6.7	S13
Figure S5. PAGE gel of TBP 6.6 and TBP 6.7	S14
Figure S6. Fold-change in SHAPE activity in the HIV-1 5'-UTR with TBP 6.6 or 6.7	S14
Figure S7. mFold predicted structures for TAR hairpin and mutants from this work.	S15
References	S16

Materials and Instrumentation

Materials

All chemicals obtained from Sigma-Aldrich unless specified
LB Miller Broth, Fisher
5-α chemically competent *E. coli*, NEB
BL21 (DE3) chemically competent *E. coli*, NEB
Agar, Fisher
Carbenicillin, GoldBio Technology
Restriction Enzymes, NEB
Isopropyl-b-D-1-thiogalactopyranoside (IPTG), GoldBio Technology
Snakeskin Dialysis Tubing 10K MWCO, Thermo Scientific
Quick Ligation Kit, NEB
Q5 High Fidelity DNA Polymerase, NEB
Oligonucleotides, IDT
Miniprep Kit, OMEGA
PageRuler Prestained Protein Ladder, Thermo Scientific
10% Ready Gel precast gels, Biorad
cOmplete ULTRA Tablets, Roche
BirA-Biotin Ligase Kit, Avidity
Casamino Acids, Fisher
Yeast Nitrogen Base without Amino Acids, BD
Yeast Extract, Alfa Aesar
Dextrose, Fisher
D-Galactose, GoldBio Technology
Peptone, Fisher
Penicillin-Streptomycin, Fisher
FITC-Conjugated anti-*myc* antibody, Abcam
Cyanine-5 labelled TAR RNA, Integrated DNA Technologies
Gene Pulser/Micro Pulser Cuvettes 2 mm, Biorad
Streptavidin-Coated Plates, Pierce
Odessey Blocking Buffer, Li-Cor
TMB One Substrate, Promega
HRP-Conjugated anti-DDDDK antibody, Abcam
Yeast Plasmid Miniprep II, Zymoprep
All water was obtained from a Milli-Q water purification system.

Instrumentation

Sonifer W-350 cell disruptor, Branson
MoFlo Flow Cytometer with a solid state iCyt 488 nm laser, Beckman Coulter
J2-21 centrifuge, Beckman Coulter
MJ mini gradient thermal cycler, Biorad
Molecular imager gel doc XR+ system, Biorad
Innova 42/42R incubator shakers, New Brunswick Scientific
NanoDrop 2000 UV-Vis Spectrophotometer, Thermo Scientific
Biacore 2000, Biacore Inc., now GE Healthcare

Protein library preparation.
EBY100 yeast (trp-, leu-, with the Aga1p gene stably integrated) and the pCTCON2 plasmid were generously provided by the Wittrup lab (MIT). The gene coding for U1A E19S was PCR amplified using primers Fwd U1A NheI (5′-ATA TAG CTA GCA TGG CCC AGG TGC AGC-3′) and Rev U1A BamHI (5′-CGG GAT CCT GCG GCC GCA ACC-3′) and cloned into pCTCon2 in frame with Aga2, an N-terminal HA-tag, and a C-terminal *myc* tag, using highlighted NheI and BamHI cloning sites. When analyzed for display, U1A E19S displayed with reasonable efficiency (at least 20-30%) and demonstrated RNA binding.

β2- β3 Loop Library Construction
To construct the β2- β3 loop library, first the BsaI restriction site was removed from the *ampR* gene of the pCTcon2 via site-directed mutagenesis using primers Fwd BsaI out (5′-CAA GGA GGT GTC GAG C GCC ACC AAC-3′) and Rev BsaI out (5′-CTC GAC ACC TCC TTG AAG ATG ACA AAA GCT TGG CC-3′). Then, this plasmid was digested with NheI and BamHI (NEB), treated with Calf Intestine Phosphatase (CIP), and extracted from a 1% agarose gel to produce a pure vector suitable for cloning. An insert encoding, from 5′ to 3′, an NheI cut site, a BsaI resatriction site, and a 3′ portion of U1A E19S from Arg52 to the C-terminus was prepared via PCR using primers Fwd U1A w BsaI (5′- ATA TAG CTA GCA GCT AGC TAG CTA GAT GGT CTC AGG GGC CAA GCT TTT GTC ATC TTC AAG GAG GTT TCG-3′) and Rev U1A BamHI (5′-CGG GAT CCT GCG GCC GCA ACC-3′). The resulting amplicon was digested with NheI and BamHI. The insert and vector were then combined using a Quick Ligation Kit (NEB) and transformed into 5-α *E.coli* (NEB) using standard methods to generate the first generation library.

Next, the β2- β3 library was created amplifying the 5′ portion of U1A E19S (Positions 1-53) with 5 sites in the β2- β3 loop region (Ser46, Ser48, Leu49, Lys50, and Met51) substituted with NNK codons. An amplicon was generated via PCR using primers Fwd U1A NheI (5′-ATA TAG CTA GCA TGG CCC AGG TGC AGC-3′) and Rev b2- b3 lib (5′-TAT AT GGTCTC GCC CCT MNN MNN MNN MNN CCG MNN TAC CAG GAT ATC CAG GAT CTG GCC-3′). To receive this insert, first generation library in pCTCON2 was also digested with NheI and BsaI, treated with CIP, and extracted from a 1 % agarose gel to give pure vector suitable for cloning. The insert and vector were then combined using a Quick Ligation Kit (NEB) according to the manufacturer's instructions. Ligated vector was purified by phenol chloroform extraction (3x), chloroform extraction (2x), and ethanol precipitation. The resulting DNA was used as a template for a second PCR with homologous recombination primers, Fwd: (5′-CTC TGG TGG AGG GCG TAG CGG AGG CGG AGG GTC GGC TAG C-3′) and Rev: (5′-CGA GCT ATT ACA AGT CCT CTT CAG AAA TCA GCT TTT GTT CGG ATC C-3′), which are designed to create an insert with ~40 base pairs of overlap with the pCTCON2 vector. The resulting amplicon, containing the randomized β2-β3 loop sequences, was then cloned into pCTcon2 using homologous recombination in EBY100 yeast.

Yeast Display Screening

Approximately 2 μg of pCTCON2 vector cut with BamH1 and Nhe1 was mixed with ~5 *μ*g of the amplified library, ethanol precipitated, and transformed via electroporation into 50 *μ*L of electrocompetent EBY100 using 2 mm cuvettes, and immediately rescued with 1 mL pre-warmed YPD for 2 hours at 30 °C. After rescue, yeast was centrifuged at 1,300 g for 1 minute, and supernatant YPD was removed. Yeast were suspended in 50 mL fresh SD-CAA (5.4 g/L Na_2HPO_4, 8.6 g/L $NaH_2PO_4 \cdot H_2O$, 20 g/L dextrose, 6.7 g/L yeast nitrogen base lacking amino acids, 5 g/L casamino acids, 200 kU/L penicillin, 0.1 g/L streptomycin).

A small portion was plated by serial dilution onto SD-CAA plates, and incubated at 30 °C for 3 days in order to determine the transformation efficiency. It was determined that an acceptable number of yeast (~10^7) were transformed. After 2-3 days of growth in SD-CAA, the library was subcultured in SD-CAA at an initial density of 0.5 x 10^7 cells/mL and grown to a density of 2.0 x 10^7 cells/mL. Yeast were subsequently subcultured in SG-CAA (Galactose containing induction media) to a concentration of 1.0 x 10^7 cells/mL and grown for 1-2 days shaking at 250 RPM at a temperature of 30 °C.

For each round of screening, approximately 10^8 cells were pelleted and washed with 500 μL of 4 °C PBS-BSA. Yeast were subsequently incubated with Cyanine-5 labeled TAR RNA, FITC-conjugated anti-myc antibody, and competitor tRNAs under the conditions given in the table below. After incubation, a final wash with ice-cold PBS-BSA was performed, and yeast positive for both FITC (display) and Cy-5 (RNA-binding) were sorted into 5 mL of SD-CAA media using a MoFlo Flow Cytometer (Beckman-Coulter). Sorted yeast were transferred to 50 mL of pre-warmed SD-CAA and incubated at 30 °C for 3 days shaking at 250 RPM. Additionally, plasmid DNA was recovered from the sorted library using a Zymoprep yeast plasmid miniprep II kit. This DNA was used to transform NEB 5α *E. coli*. Additionally, following the third round, this DNA was used as template for a diversified library. The screening is summarized in **Table S1.**

Table S1. Experimental conditions for yeast display

	R1	R2	R3	R4	R5	R6
[TAR-Cy5, nM]	10,000	1,000	500	500	100	10
[tRNA, nM]	4,000	4,000	50,000	50,000	50,000	50,000
incubation time	1 hour	30 min.	30 min.	30 min.	30 min.	30 min.
incubation temp	25 °C	25 °C	37 °C	37 °C	37 °C	37 °C
yeast screened	~ 4 x 10^7	~6 x 10^7	~3 x 10^7	~3 x 10^7	~3 x 10^7	~3 x 10^5
yeast sorted	~ 1 x 10^6	~ 6 x 10^5	~2 x 10^5	4.7 x 10^5	~2 x 10^6	13.4 x 10^6

Diversification of the 2nd generation library

A modified pCTcon2 plasmid, named second generation library-pCTCON2, was constructed to receive the second-generation library as follows. A pCTCON2 plasmid containing wtU1A downstream of Aga2 and a linker was digested with restriction enzymes MluI and NotI (both from NEB), followed by treatment with Calf Intestine Phosphatase (NEB), and extracted from a 1% agarose gel to give pure vector suitable

for cloning. Then, two oligos, Fwd: (5′- CGC GTC CTA ACC ACA CTA TTT ATA TGA GAC CAC TCT AGA GGT TCC CCG GTT GC-3′) and Rev:(5′- GGC CGC AAC CGG GGA ACC TCT AGA GTG GTC TCA TAT AAA TAG TGT GGT TAG GA-3′), were treated with T4 Polynucleotide Kinase (NEB) according to the manufacturer's instructions. The phosphorylated oligos were then heated up to 94 °C for 5 minutes and allowed to room temperature slowly, over 5 minutes. The annealed oligos were then cloned into the vector using standard methods to create second generation library-pCTcon2.

The second-generation library was built by pooling the library members isolated in the third round of screening and then, within those genes, introducing diversity at the C-helix position on the U1A scaffold. To begin, yeast collected from the third round of screening were cultured and minipreped (Zymoprep) to give a pool of pCTcon2 plasmids containing U1A β2- β3 mutants. The U1A mutant genes were amplified by PCR using the primers Fwd U1A (5′- ATA TAG CTA GCA TGG CCC AGG TGC AGC-3′) and Rev c-helix lib (5′-TAT ATG GTC TCT TGG CMN NGA TMN NMN NGT CGG TGC GCG CAT ACT GGA TAC G-3′). The resulting amplicon was digested with NheI and BsaI (NEB). To receive this insert, second generation library-pCTcon2 was also digested with NheI and BsaI, then treated with Calf Intestine Phosphatase (NEB), and extracted from a 1% agarose gel to give pure vector suitable for cloning. The insert and vector were then combined using a Quick Ligation Kit (NEB) according to the manufacturer's instructions. Ligated vector was purified by phenol chloroform extraction (3x), chloroform extraction (2x), and ethanol precipitation. The resulting DNA was used as a template for a second PCR with homologous recombination primers (Fwd: 5′-CTC TGG TGG AGG GCG TAG CGG AGG CGG AGG GTC GGC TAG C-3′; Rev: 5′-CGA GCT ATT ACA AGT CCT CTT CAG AAA TCA GCT TTT GTT CGG ATC C-3′). The resulting amplicon, containing β2- β3 loop sequences isolated from three rounds of screening and a randomized c-helix region, was then cloned into pCTcon2 using homologous recombination in EBY100 yeast.

Protein Purification

After transforming the library from the sixth generation of TAR binding proteins into *E. coli*, individual library members were isolated, and the TAR binding proteins cloned to contain C-terminal His_6 and FLAG tags (DYKDDDDK) in a pET plasmid. The gene was amplified from pCTCON2 and cut using enzymes NcoI and PacI (Forward Primer : 5′-ATA TAC CAT GGC CCA GGT GCA GC, Reverse Primer: GTT AAT TAA CTA TTA CTT GTC GTC ATC GTC TTT GTA GTC GTG ATG ATG GTG ATG ATG TGC GGC CGC AAC C - 3′). These constructs were confirmed by DNA sequencing (all sequencing performed by GENEWIZ, South Plainfield, NJ), and subsequently transformed into BL21*E. coli* for protein expression. Cells were grown in 100-500 mL cultures to OD_{600} = ~0.6 and induced with 1 mM IPTG at 25 °C for 4-12 hours. Cells were then collected via centrifugation, resuspended in HEPES buffer (10 mM HEPES, 50 mM KCl, 30 mM NaCl, 1 mM $MgCl_2$, 1 mM EDTA, pH = 7.4) with protease inhibitor tablets, and stored at -20 °C. Frozen cell suspensions were thawed, then sonicated for 2 minutes. The lysate was cleared via centrifugation (9000 RPM, 20 minutes) and the supernatant was mixed with 400 μL of Ni-NTA agarose resin for 10 minutes. The resin was collected by

centrifugation (4750 RPM, 5 minutes). The resin was then washed with 30 mL of HEPES buffer containing 20 mM imidazole, followed by 10 mL of HEPES buffer containing 50 mM imidazole. Proteins were then eluted using 4 mL of HEPES buffer containing 400 mM imidazole.

Eluted proteins were dialyzed in snakeskin tubing against 2 L of HEPES buffer, and subsequently against 2 L of phosphate buffer (20 mM phosphate, 150 mM NaCl, pH = 7.4). Following dialysis, protein was analyzed for purity via SDS – PAGE. Purified proteins were quantified using absorbance at a wavelength of 280 nM, using calculated extinction coefficients.

Enzyme-Linked Immunosorbent Assay (ELISA)

ELISA assays were performed using clear, streptavidin coated, 96-well plates (Pierce). The plate was pre-incubated for 1 hour with wash buffer (20 mM phosphate, 150 mM NaCl, 0.05% Tween-20 and 0.1 mg/mL BSA, pH = 7.4). Following pre-incubation, 100 μL of TAR (5′–GGC AGA UCU GAG CCU GGG AGC UCU CUG CC-3′) or U1hpII (5′ – AGC UUA UCC AUU GCA CUC CGG AUG AGC U – 3′RNA modified with a 5′ biotin (purchased from Integrated DNA Technologies, Coralville IA) was incubated at a concentration of 50 nM for 2 hours. Wells were washed 3 times with 200 μL of wash buffer shaking for 5 minutes. Subsequently, 100 μL of buffer containing 20 nM of either a TAR binding protein or wild-type U1A containing a C – Terminal FLAG-Tag was incubated for 1 hour, and washed three times. Following this, a 1:10,000 dilution of HRP-conjugated anti-FLAG antibody (Abcam, ab2493) was incubated in 100 μL of Odyssey Blocking Buffer (Li-Cor) for 30 minutes at 25 °C, and washed 4 times. Colorimetry was developed for 30 min. using 100 μL of TMB-One substrate. Absorbance was measured at 655 nM on a plate reader.

TAR binding proteins 6 and 7, from the **6**th generation of screening (TBP 6.6 and 6.7) demonstrated exceptional affinity for TAR and low affinity for U1hpII, and were further analyzed quantitatively. The ELISA protocol remained the same, but concentrations of the TAR binding protein were varied, in triplicate, from 0.05 nM to 200 nM. This allowed a binding curve to be constructed, and a binding constant estimated. Both proteins analyzed in this manner were estimated to bind TAR with a K_D of ~5 nM, and were sent to be analyzed via SPR.

To determine which portions of TAR are most important for binding, variants of TAR RNA were ordered, one missing TAR's characteristic stem-bulge (HP1: 5′ - GGC AGA GAG CCU GGG AGC UCU CUG CC - 3′), three which substituted two bases on the 6-base loop for their reverse complements (HP2: 5′-GGC AGA UCU GAG CGA GGG AGC UCU CUG CC - 3′, HP3: 5′ - GGC AGA UCU GAG CCU CCG AGC UCU CUG CC - 3′, and HP4: 5′ - GGC AGA UCU GAG CCU GGC UGC UCU CUG CC-3′), two which replaced G-C Watson-Crick base pairs on the stem with A-U base pairs (HP5: 5′-GGC AGA UCU GAG UCU GGG AAC UCU CUG CC-3′ and HP6: 5′– GGC AGA UCU GAA UCU GGG AAU UCU CUG CC-3′), and finally, a variant of TAR found in the Bovine Immunodeficiency Virus (BIV) (BIV TAR: 5′–GGC UCG UGU AGC UCA UUA

GCU CCG AGC C-3′). The ELISA protocol described above, using 20 nM TBP 6.6 and 6.7, was used with the TAR variants in place of the original TAR.

Surface Plasmon Resonance

TBP 6.6 was sent to be analyzed by SPR with the same construct as was used in ELISA. A version of TBP 6.7 was prepared which did not have a FLAG-Tag. The gene was PCR amplified from its pET plasmid (FP: 5´A-TA TAC CAT GGC CCA GGT GCA GC-3´, RP: 5´-GTT AAT TAA CTA TTA GTG ATG ATG GTG ATG ATG TGC GGC CGC AAC C-3´), cut with NcoI and PacI, ligated into pET plasmid, and purified as shown above.

A streptavidin-coated sensor chip (Sensor chip SA, GE Healthcare) was used to coat 25 response units (RU) of 5´-biotinylated U1hpII (5´-AGC UUA UCC AUU GCA CUC CGG AUG AGC U-3´) on flow cell 1, and 25 RU of TAR (5´-GGC AGA UCU GAG CCU GGG AGC UCU CUG CC-3´) on flow cell 3, leaving flow cells 2 and 4 blank for background correction. Proteins were serially diluted in running buffer (10 mM Tris-HCl, pH 8, 150 mM NaCl, 5% glycerol, 62.6 μg mL^{-1} bovine serum albumin, 1 mM dithiothreitol, 0.05% surfactant P20, and 125 μg mL^{-1} yeast tRNA) to the concentrations indicated (Fig. 4D) and injected at 20 °C with a flow rate of 50 uL/ min over all surfaces consecutively. In each for three independent experiments, triplicate injections were fully randomized and interspersed with buffer injections to allow double referencing.[1] After each protein injection the surface was regenerated with a 1-min 2 M NaCl injection, followed by a buffer injection. Data was processed using Scrubber and analyzed using CLAMP XP[2] and a simple 1:1 Langmuir interaction model with a correction for mass transport. Association and dissociation rates are listed in **Table S2**. U1A RRM1 was injected for comparison, using the conditions described above.

Table S2. Kinetic Data from SPR experiments (shown in Figure 4D of the paper)

TBP 6.6 and TAR RNA

	experiment 1	experiment 2	experiment 3	average	SEM	SD	avgerage with SEM
k_a	1.16E+07	1.23E+07	1.42E+07	1.27E+07	7.58E+05	1.31E+06	1.3 ± 0.1E+07
k_d	1.91E-02	1.38E-02	1.39E-02	1.56E-02	1.76E-03	3.04E-03	1.6 ± 0.2E-02
K_D	1.65E-09	1.12E-09	9.79E-10	1.25E-09	2.03E-10	3.51489E-10	1.3 ± 0.2 nM

TBP 6.7 and TAR RNA

	experiment 1	experiment 2	experiment 3	average	SEM	SD	avgerage with SEM
k_a	1.81E+07	4.64E+07	2.32E+07	2.92E+07	8.72E+06	1.51E+07	5.6 ± 0.7E+06
k_d	9.01E-03	2.31E-02	1.30E-02	1.50E-02	4.20E-03	7.28E-03	8.0 ± 0.9E-03
K_D	4.98E-10	4.98E-10	5.60E-10	5.19E-10	2.06E-11	3.55991E-11	1.5 ± 0.3 nM

U1A and U1hpII RNA

	experiment 1	experiment 2	experiment 3	average	SEM	SD	avgerage with SEM
k_a	1.21E+07	7.13E+06	1.38E+07	1.10E+07	2.00E+06	3.47E+06	1.1 ± 0.2E+07
k_d	5.41E-04	3.60E-04	3.64E-04	4.22E-04	5.97E-05	1.03E-04	4.2 ± 0.6E-04
K_D	4.47E-11	5.05E-11	2.64E-11	4.05E-11	7.27E-12	1.25876E-11	0.041 ± 0.007 nM

SEM = standard error of the mean
SD = standard deviation

Isothermal Titration Calorimetry.
Isothermal titration calorimetry (ITC) was performed using a MicroCal iTC200 calorimeter maintained at 25 °C. TBP 6.6 and TBP 6.7 were expressed with C-terminal His tags, and purified by as described above. Purified proteins were dialyzed extensively in phosphate buffer (20 mM sodium phosphate pH 7.4, 150 mM NaCl). Truncated HIV Tat peptide was ordered (Genscript Corp), and resuspended in this phosphate buffer. HIV TAR RNA was placed in the sample cell at concentrations ~6 μM, and ~60 μM Tat peptide was injected in 2.49 μL increments (16 injections total), with an initial injection of 0.4 μL, at 180 second intervals using a stirring speed of 750 rpm. Displacement experiments were performed by titrating 65.3 μM tat into a pre-formed 1:1 complex of TAR and TBP 6.7 (6.1 μM each), or a pre-formed 1:2 complex of TAR and TBP 6.6 (6.1 μM TAR and 13.8 μM TBP 6.6) Data were analyzed using Origin7.0 (MicroCal, ITC200) using a one set of sites binding model for fitting. All data were reference subtracted by subtracting the mean heat of dilution from each data point.

RNA preparation for SHAPE. 362 nt long NL4-3 5' UTR RNA was prepared by *in vitro* transcription using the MegaShortScript kit (Ambion/Life Technologies) according to manufacturers' recommendations. DNA template used in the transcription reaction was generated by PCR from a proviral pNL4-3 plasmid using high fidelity platinum Taq DNA polymerase (Invitrogen) and forward and reverse primers "T7L" (5′-TAATACGACTCACTATAGGTCTCTCTG-3′) and "369R" (5'-GCTTAATACCGACGCTCTCGC-3′) respectively. The forward primer was designed to introduce T7 promoter sequence at the 5' end of the 5' UTR. The RNA was then treated with Turbo DNase I for 1 hour at 37 °C, heated at 85 °C for 2 min and run on a denaturing gel (5% polyacrylamide (19:1)), 1x TBE, 7M urea) at constant temperature (45°C, 30W max). The 5' UTR band was then excised, electro-eluted at 200 V for 2 hours at 4 °C, ethanol precipitated and stored at -20 °C in TE buffer (10 mM Tris, pH 7.6; 0.1 mM EDTA) prior to use.

SHAPE and RNA migration experiments.
In 5 different tubes, 40 pmoles of RNA in a total volume of 10 μl was refolded by heating to 85 °C for 2 minutes, followed by slow cooling to 25 °C for 15 minutes (ramp rate 0.1°C/sec). Meanwhile, two-fold serial dilutions of U1A mutant protein (20 picomoles/μl, 10 picomoles /μl , 5 picomoles /μl , 2.5 picomoles /μl) were made in the protein storage buffer (20 mM phosphate, 150 mM NaCl at pH 7.4 containing 10% glycerol). The volume in each tube was brought to 284 μl by adding 274 μl of nuclease free water (Invitrogen). 16 μl of each protein dilution or 16 μl of protein storage buffer alone was incubated with the folded RNA at 37 °C for 10 mins. 144 μl of each RNA-protein mixture was aliquoted into two tubes labeled as "NMIA+" and "NMIA-". RNA in the "NMIA+" tubes was chemically modified by incubation with 16 μl of 30 nM NMIA in anhydrous DMSO at 37 °C for 20 minutes. To the "NMIA-" tubes, 16 μl of anhydrous DMSO was added and these tubes were also incubated at 37 °C for 20 minutes. Protein in both the "NMIA+" and "NMIA-" tubes was removed by phenol-chloroform extraction. For this, 140 μl of water followed by 300 μl of phenol:chloroform:iso-amyl alcohol mixture pH 6.8 (Ambion) was added to each tube and spun at 14000 rpm at 4 °C for 5 mins. 250 μl of

the aqueous phase was recovered and ethanol precipitated. The RNA pellet was suspended in 12 μl nuclease free water . 3 picomoles of each RNA were then used to generate cDNA library for each RNA. Subsequent cDNA processing/fractionation and SHAPE data analysis were conducted as previously described.[3]

To check the homogeneity of the RNA samples, 20 μl of the SHAPE reaction mix containing 16 μl of protein storage buffer alone was sampled out just before the addition of NMIA and fractionated on a native gel [4% polyacrylamide (19:1)), 1x TBE] at constant voltage of 200V at 4 °C for 5 and half hours. RNA bands were visualized by SYBR Green II RNA Gel Stain (Life Technologies)

Inhibition of Tat/TAR-dependent Transcriptional Activation

A DNA fragment (-477 to +568) containing the HIV 5′ – LTR was PCR amplified from the plasmid pLAI.BS (FP: TCTAGAACTAGTGGATCTTAG, RP: GCTACAACCATCCCTTCAGAC). An *in vitro* transcription reaction was performed in a 40 microL reaction containing 18 microL of HeLa nuclear extract in 20 mM HEPES, 80 mM KCL 3 mM $MgCl_2$, 2 mM DTT 10 microM $ZnCl_2$, 15 U rRNasin, 1 microg creatine kinase, 10 mM creatine phosphate, 250 microM of GTP, ATP, and CTP, 50 microM UTP, and 10 microCi [α-^{32}P]UTP. Reactions contained the PCR amplicon template, Tat Protein, and C-Terminal His_6-tagged TBP 6.7 in the concentrations given in the table below.

The reactions were incubated for 5 hours at 37 °C, and quenched by addition of 200 microliters HSCB buffer (25 mM Tris-HCL, 400 mM NaCl, 0.1% SDS, pH = 7.5). Following reaction stop 60 microg of glycogen was added to each reaction as a carrier, as well as 1 microL of a 60 base radio-labelled RNA was added as a loading control. Proteins were extracted using phenol/chloroform/isoamyl alcohol, and nucleic acids were ethanol precipitated. Ethanol precipitated nucleic acids were resuspended in RNA loading dye, and seperated via PAGE. PAGE gels were then developed using a phosphor imaging screen, and a Typhoon imager.

A ~500 bp transcript was quantified using ImageQuant software from GE Healthcare. The 60 base spikant band was used to confirm that no significant variations occurred during the Phenol-Chloroform extraction and ethanol precipitation.

Reaction	[Template]	[Tat]	[TBP 6.7]
1	10 nM	2 μM	–
2	10 nM	2 μM	2 μM
3	10 nM	2 μM	0.2 μM

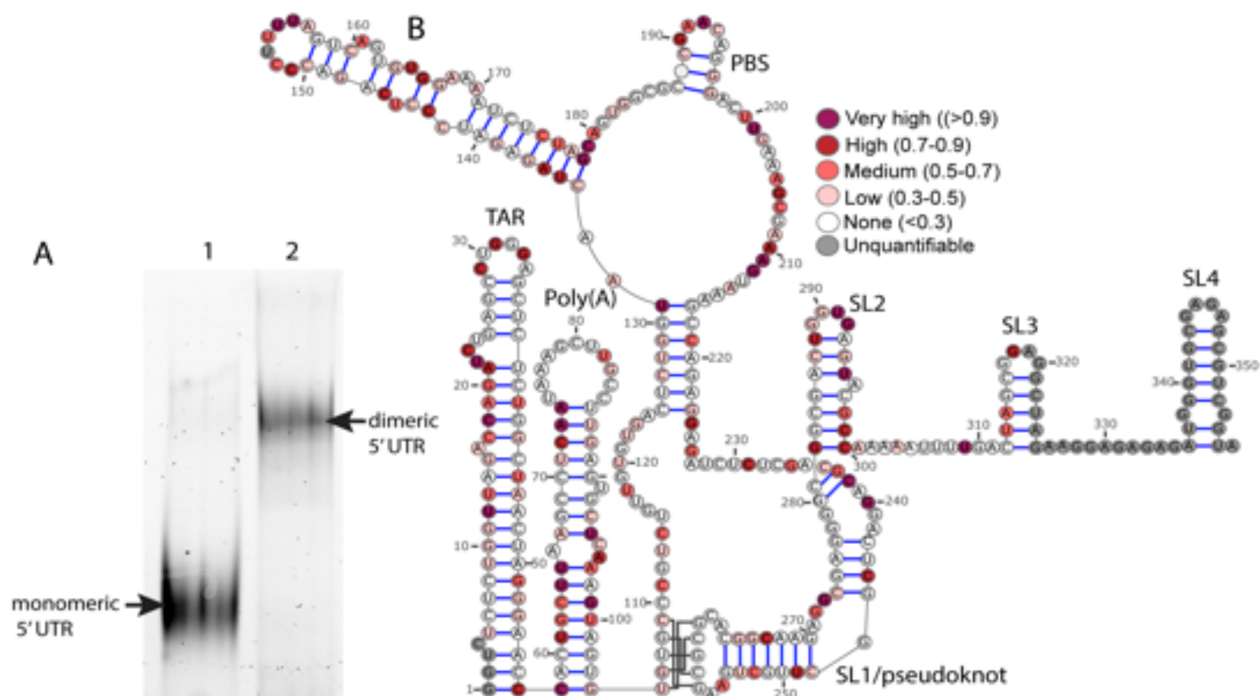
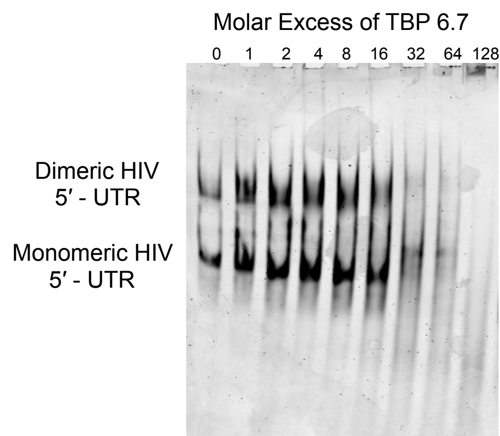


Figure S1: (A) RNA folding conditions that allow NL4-3 5' UTR RNA to adopt a monomeric or dimeric configuration. A sample (20µl) of the folded NL4-3 5' UTR RNA from the SHAPE experiments, immediately prior to the addition of the electrophile NMIA, was fractionated on a 4% native gel for 5.5 hours at 4°C at constant 200V. RNA in lane 1 was folded in the absence of any salts whereas that in lane 2 was folded in the presence of 80mM Tris/HCl pH 7.5, 280mM KCl, 0.4mM EDTA, 8mM $MgCl_2$. (B) SHAPE reactivity of NL4-3 5' UTR RNA, color-coded on the proposed pseudo-knot 5' UTR monomeric structure.

TBP 6.7 Binding to Full HIV 5′ - UTR



TBP 6.7 Binding to TAR – Poly(A)

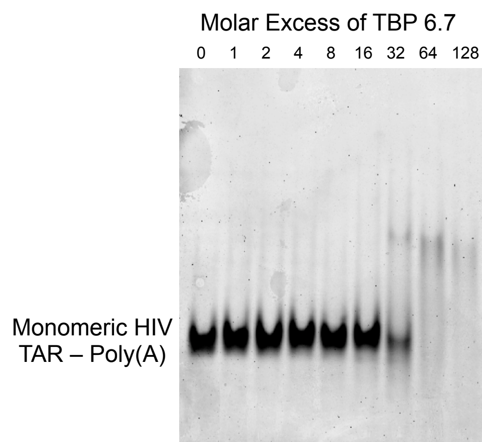


Figure S2. In both the full HIV 5′-UTR and the isolated TAR–Poly(A) element, appreciable shifts are seen with a 32-fold excess of TBP 6.7. This shift occurs in the full 5′ - UTR in both the monomeric and dimeric forms. It is clear that this is not degradation, as all shifts appear to involve *increasing* mass. This qualitative data shows binding occurring in the absence of a solid support.

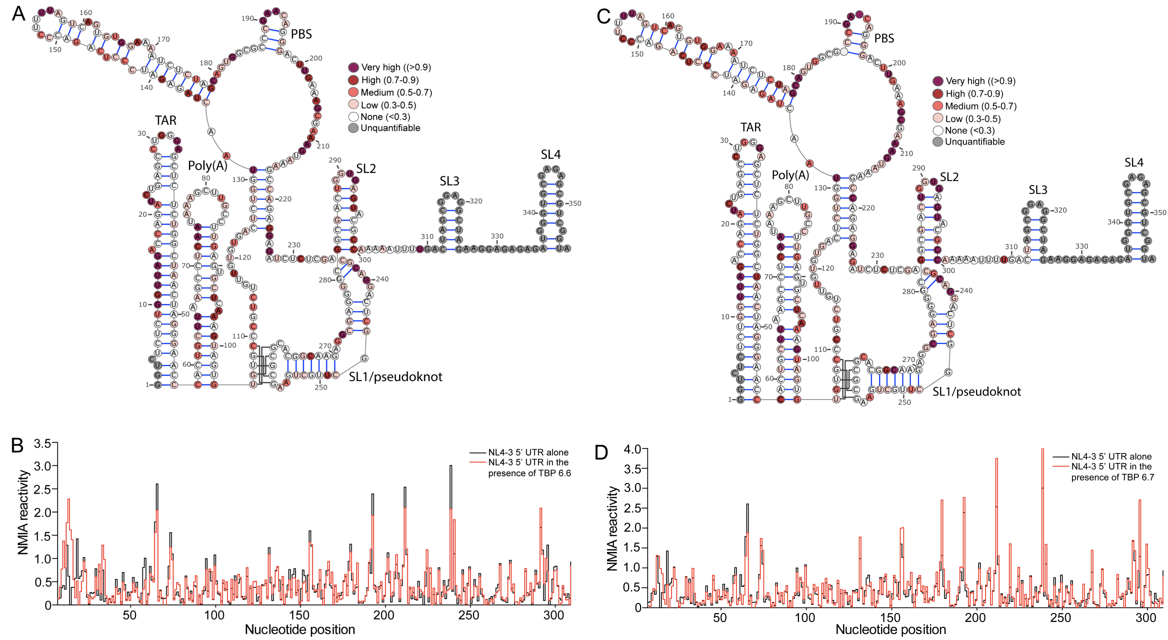


Figure S3 and S4: SHAPE analysis of the monomeric NL4-3 5' UTR RNA in the presence of 4X molar of TBP 6.6. and TBP 6.7 (A) SHAPE reactivity of 5' UTR obtained in the presence of TBP 6.6, color-coded on the proposed pseudo-knot 5' UTR monomeric structure. (B) Step plots comparing the normalized NMIA reactivity of the 5' UTR in the presence (red) or absence (red) of TBP 6.6. (C) SHAPE reactivity of 5' UTR obtained in the presence of TBP 6.7, color-coded on the proposed pseudo-knot 5' UTR monomeric structure. (D) Step plots comparing the normalized NMIA reactivity of the 5' UTR in the presence (red) or absence (red) **of** TBP 6.7**.**

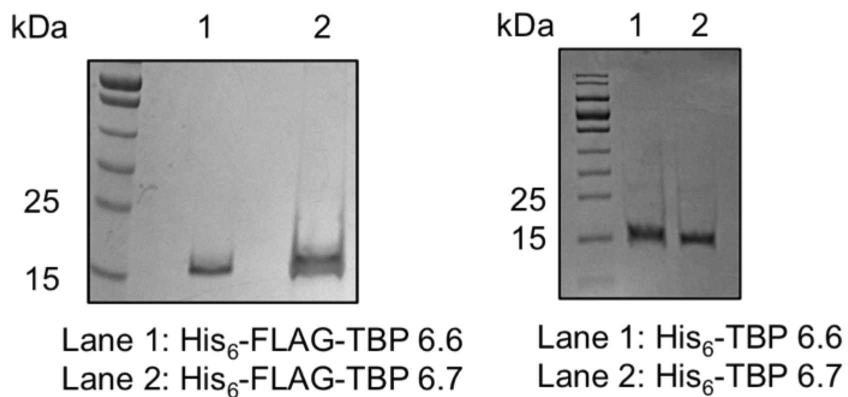


Figure S5. Coommassie stained PAGE gel of all proteins used in this work

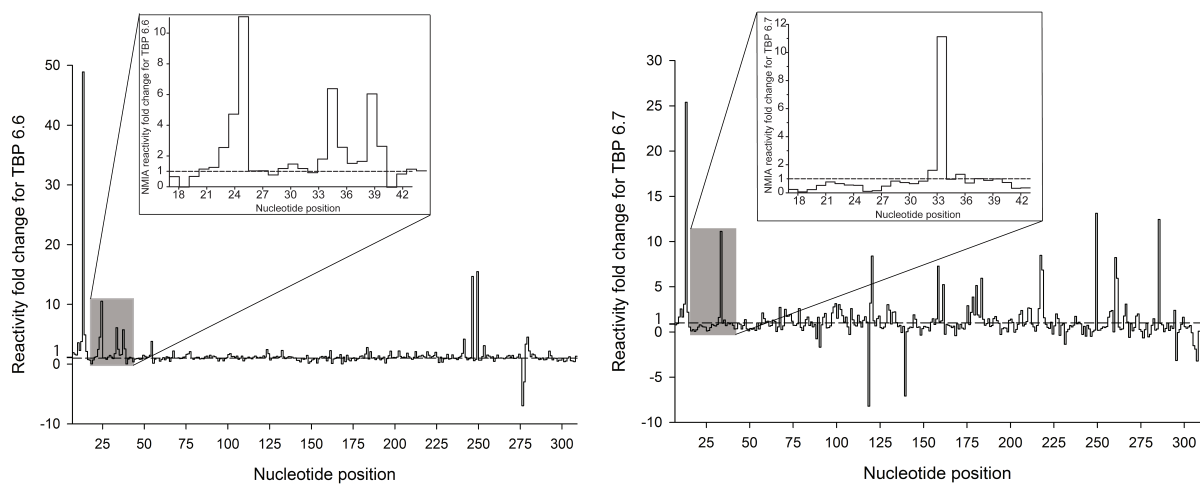
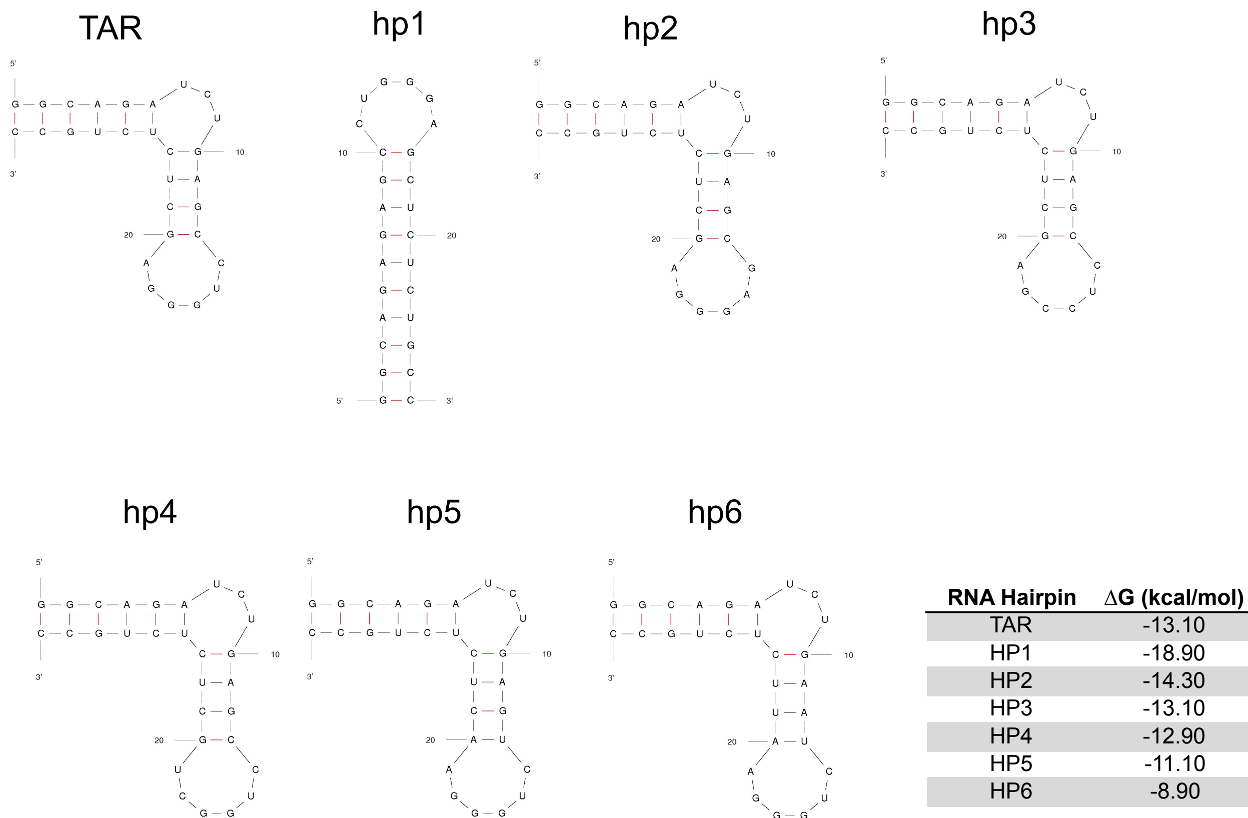


Figure S6. Fold-change in SHAPE reactivity over the HIV-1 5’-UTR upon addition of TBP 6.6 or 6.7



RNA Hairpin	ΔG (kcal/mol)
TAR	-13.10
HP1	-18.90
HP2	-14.30
HP3	-13.10
HP4	-12.90
HP5	-11.10
HP6	-8.90

Figure S7. mFold predicted structures and folding energies of TAR and mutants used in this work.

References.

[1] Myszka, D. G. (1999) Improving biosensor analysis, *Journal of molecular recognition : JMR 12*, 279-284.
[2] Myszka, D. G., and Morton, T. A. (1998) CLAMP: a biosensor kinetic data analysis program, *Trends in biochemical sciences 23*, 149-150.
[3] Lusvarghi, S., Sztuba-Solinska, J., Purzycka, K. J., Rausch, J. W., and Le Grice, S. F. (2013) RNA secondary structure prediction using high-throughput SHAPE, *Journal of visualized experiments : JoVE*, e50243.