

Supporting Information

for

Nucleocapsid Protein Preferentially Binds the Stem-loop of Duplex/Quadruplex Hybrid that Unfolds the Quadruplex Structure

Content

Experimental details

- Figure S1.** Sequence information of NCp7 and LTR-III
- Figure S2.** LTR-III forms a stable duplex/quadruplex hybrid structure.
- Figure S3.** Interaction of NCp7 with LTR-III
- Figure S4.** Fluorescence spectra of ThT in the absence and presence of LTR-III
- Figure S5.** CD analysis of the unfolding of LTR-III upon NCp7 binding
- Figure S6.** FRET analysis of the unfolding of LTR-III upon NCp7 binding in the presence of the complementary sequence
- Figure S7.** Hoogsteen hydrogen bond distance plots of G-tetrad in free LTR-III
- Figure S8.** Hoogsteen hydrogen bond of G-tetrad of LTR-III with the interaction of NCp7
- Figure S9.** The ITC binding isotherm of LTR-III and LTR-III mutant with NCp7
- Figure S10.** The region of NCp7 that involved in the interaction with LTR-III
- Figure S11.** RMSF analysis of residues of NCp7 when interacting with LTR-III over MD simulation
- Figure S12.** Binding mode of LTR-III in cartoon representation and NCp7 in electrostatic potential representation on the surface.
- Figure S13.** The interaction of LTR-III with NCp7 in cartoon representation.
- Figure S14.** NCp7 unfolds the G4 structure of LTR-III RNA

Experimental details

1. Materials

The protein was expressed and expressed as describing previously.¹ Oligonucleotides were purchased from Sangon Biotech, China (Shanghai, China), and were used without further purification. Samples were dissolved in ultrapure water, and then dialyzed overnight against 50 mM potassium phosphate buffer (pH 7.0). Concentrations were determined using UV spectroscopy by recording the absorbance at 260 nm. Oligonucleotide extinction coefficients (ϵ_{260}) were calculated on the website of <https://www.atdbio.com/tools/oligo-calculator>. All samples were annealed by heating at 95°C for 5 min and then gradually cooled to room temperature to induce G4 formation.

2. Electrophoretic Mobility Shift Assay (EMSA)

EMSA were performed with 5 μ M LTR-III with increasing concentrations of NCp7 in binding buffer (50 mM Tris, 100 mM KCl, pH 7.0). Samples were loaded on 12.5% native polyacrylamide gels (29:1 acrylamide:bis acrylamide ratio) in 0.5 x Tris-Boric acid buffer (54 g Tris, 27.5 g Boric acid) at 100 V, 4°C for 1 h. Gels were stained with gel red nucleic acid dye (Thermo Fisher) for 10 min, and then photographed with a gel imager in UV mode.

3. ThT assay with fluorescence competition experiment

The fluorescence measurements were performed on a Hitachi F-4600 fluorescence spectrophotometer equipped with a temperature controller. Samples were placed in a capped 1 cm path length cuvette and the excitation wavelength was set at 420 nm by scanning the emission spectra between 430 nm and 700 nm. The slit width for excitation and emission spectra was 5 and 2.5 nm, respectively. Spectra in ThT assay were recorded on 10 μ M ThT and 100 μ M LTR-III DNA after incubation with different concentrations of NCp7 for 10 min at 25°C in potassium phosphate buffer (50 mM, pH 7.0). For the ThT assay on RNA, NCp7 was titrated into 100 μ M LTR-III RNA containing 10 μ M ThT (in 20 mM Tris-buffer, 140 mM KCl, pH 7.4). The samples were incubated for 30 min at 37°C and the spectra were recorded with the excitation of 420 nm.

4. Fluorescence resonance energy transfer (FRET) experiments

Fluorescence spectra were recorded on a F-4600 FL Spectrophotometer by using 50 μ l quartz cuvettes. Different concentrations of NCp7 were incubated with 600 nM fluorophores labeled LTR-III (F-LTR-III-T) in potassium phosphate buffer (50 mM, pH 7.0) at 25°C for 30 min in dark prior to fluorescence measurements. The spectra were recorded with excitation at 470 nm.

To analyze the effect of complementary sequence (CS) on the NCp7 mediated unfolding of LTR-III, the FRET analysis was conducted on F-LTR-III-T with the addition of different concentrations of NCp7 in the presence or absence of CS. The assay was performed in 20 mM tris-buffer containing 140 mM KCl (pH 7.4), and the fluorescence spectra were recorded after incubation in the dark for 1 h at 25°C.

5. Isothermal Titration Calorimetry (ITC).

ITC measurements were performed in a VP-ITC titration calorimeter. The solution samples were thoroughly degassed before loading. The reference cell was filled with degassed buffer. The NCp7 (40 μ M) was placed in the sample cell, and the LTR-III solution (120 μ M, 70 μ L) in the same buffer was filled in syringe. The titration was performed in 2 μ L of each injection (20 injections in total) at 2 min intervals at 25°C.

6. Circular Dichroism (CD)

CD measurements were performed on a Jasco J-810 CD spectrometer flushed with high purity nitrogen. Spectra were recorded from 340 to 200 nm with a scan speed of $100 \text{ nm}\cdot\text{min}^{-1}$ and a data pitch of 1 nm in a 1.0 mm path-length quartz cuvette (cleaned by distilled water and dried with nitrogen gas). A band width of 1 nm was used with a detector response time of 1 s. LTR-III were prepared to a final concentration of $5 \mu\text{M}$ in potassium phosphate buffer (50 mM, pH 7.0). The spectrum of buffer was recorded for reference. All measurements were repeated three times. For the RNA interaction the CD spectra were recorded on $5 \mu\text{M}$ LTR-III RNA after incubation with different concentration of NCp7 at 37°C for 1 h (in 20 mM Tris-buffer, 140 mM KCl, pH 7.4).

7. NMR Spectroscopy

NMR spectra were recorded on a Bruker Avance 850 MHz spectrometer equipped with TCI CryoProbe. 2D ^1H - ^{15}N HSQC spectra were recorded on ^{15}N -labeled NCp7 proteins in potassium phosphate buffer (50 mM, pH 7.0), and ^1H NMR spectra were collected on $200 \mu\text{M}$ LTR-III at 25°C in 50 mM potassium phosphate buffer (pH 7.0). Data were processed and analyzed using Sparky or TopSpin.

8. Electrospray ionization mass spectroscopy (ESI-MS)

The ESI-MS spectra were recorded on a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) equipped with a nanoESI source. Samples were collected from size-exclusion chromatography and were directly infused for detection. Data were processed using XCalibur software (version 2.0, Thermo Finnigan).

9. Constructing structural model for protein and DNA

The LTR-III and NCp7 motifs in the simulation were from 6H1K² and 2EXF³, respectively. Their initial binding mode was referred to 2EXF, in which the π - π stacking among the aromatic rings on Phe16, Trp37 from NCp7 and T6, G7 from LTR-III likely account for critical interactions, and these interactions pull T6 and G7 out of the LTR-III base pair structure.³ Given the missing signals of G5 and G6 bases in our system, it is plausible that G5 and G6 bases would form similar interactions with Phe16 and Trp37 as that in 2EXF. The initial binding mode was constructed accordingly.

10. Force field Modeling

All the hydrogen atoms in the system were added by GROMACS package⁴. The force field of the two Zn ions and the zinc-binding residues was parameterized by using "MCPB.py" model⁵, in which the three cysteine residues and one histidine residue were treated as covalently bonded with the zinc center, thus forming a tetrahedral structure. The parameters for the rest of the system were retrieved from ff99sb Amber force field.⁶ The Amber99sb force field⁷ and its modified versions⁸ have been widely applied to study the stability of G-Quadruplex DNA structures. The LTR-III-NCp7 complex was soaked into a cubic box of 13405 classical TIP3P water molecules,⁹ with a minimum distance of 12 Å to the box edge under periodic boundary conditions (PBC). 20 positive counter ions (K^+) were added to neutralize the charge of the system, among which, to stabilize the G-quadruplex structures, two K ions were intentionally set in the central channel between each pair of quadruplex. To further study the role of NCp7 during the binding, a model of free LTR-III dissolved in a cubic box of 9513 TIP3P water molecules and 25 positive counter ions (K^+) was also simulated as comparison.

11. Molecular dynamics (MD) simulations

The MD simulations were performed by using Gromacs 5.1 software package.⁴ For both free LTR-III and LTR-III-NCp7 systems, the first energy minimization step was performed with steepest descent algorithm for about 20,000 cycles. Then a 350 ns trajectory with the time step size at 2 fs of NPT ensemble was produced for analysis, whose temperature (300 K) and pressure (1 bar) were maintained using Berendsen temperature and pressure coupling method.¹⁰ The same thermal bath has been applied to study Guanine Quadruplex folds before.¹¹ The cutoff of Lennard-Jones 12-6 potential to represent Van der Waals interactions is 10 Å, while the long-range electrostatics interactions calculated by the particle mesh Ewald (PME) method are truncated at 10 Å as well. The Lincs algorithm was performed to constrain all hydrogen bonds.¹²

12. MD Analysis

To examine the binding details between LTR-III and NCp7 and their dynamical changes, the two trajectories of 350 ns NPT MD simulations were used for analysis. The root-mean-square deviations (RMSDs), root-mean-square fluctuations (RMSFs), principal components analysis (PCA), and other analyzing methods were performed by Gromacs inbuilt tools.

References

1. Y. Sheng, K. Cao, J. Li, Z. Hou, S. Yuan, G. Huang, H. Liu and Y. Liu, *Chem Eur J*, 2018, **24**, 19146-19151.
2. E. Butovskaya, B. Heddi, B. Bakalar, S. N. Richter and A. T. Phan, *J Am Chem Soc*, 2018, **140**, 13654-13662.
3. S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. Salgado, B. P. Roques, Y. Mely, S. Bouaziz and N. Morellet, *J Mol Biol*, 2008, **383**, 1112-1128.
4. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845-854.
5. P. Li and K. M. Merz, *J Chem Inf Model*, 2016, **56**, 599-604.
6. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Structure, Function, and Bioinformatics*, 2006, **65**, 712-725.
7. (a) L. J. Aaldering, V. Poongavanam, N. Langkjaer, N. A. Murugan, P. T. Jørgensen, J. Wengel and R. N. Veedu, *Chembiochem : a European journal of chemical biology*, 2017, **18**, 755-763; (b) P. Mitrasinovic, *Croatica Chemica Acta*, 2019, **92**.
8. (a) M. Ebrahimi and T. Khayamian, *Med Chem Res*, 2014, **23**, 1327-1339; (b) A. Głuszyńska, B. Juskowiak, M. Kuta-Siejkowska, M. Hoffmann and S. Haider, *Molecules*, 2018, **23**, 1134; (c) J. Jamroskovic, M. Doimo, K. Chand, et al, *J Am Chem Soc*, 2020, **142**, 2876-2888; (d) B. Prasad, R. N. Das, J. Jamroskovic, R. Kumar, M. Hedenström, N. Sabouri and E. Chorell, *Chem Eur J*, 2020, **26**, 9561-9572; (e) R. Reshetnikov, A. Golovin, V. Spiridonova, A. Kopylov and J. Šponer, *J Chem Theory Comput*, 2010, **6**, 3003-3014.
9. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926-935.
10. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J Chem Phys*, 1984, **81**, 3684-3690.
11. B. Islam, P. Stadlbauer, S. Neidle, S. Haider and J. Sponer, *J Phys Chem B*, 2016, **120**, 2899-2912.
12. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J Comput Chem*, 1997, **18**, 1463-1472.

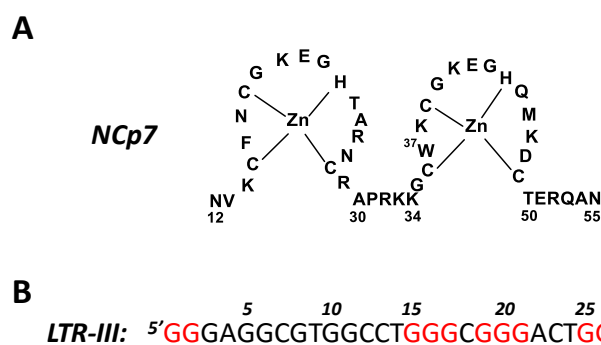


Figure S1. Sequence information of NCp7 and LTR-III. **(A)** The sequence of NCp7 (aa 12-55) and zinc binding sites. **(B)** Sequence of the LTR-III. G bases involved in G4 formation are shown in red.

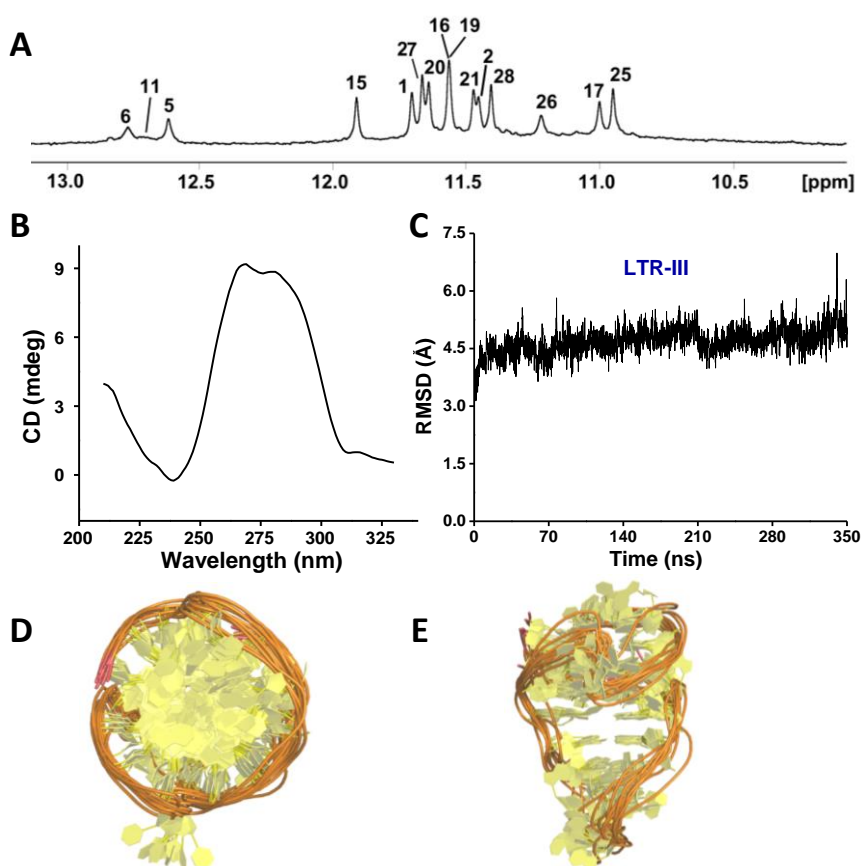


Figure S2. LTR-III forms a stable duplex/quadruplex hybrid structure. **(A)** ^1H NMR imino proton spectrum of 200 μM LTR-III at 25°C in 50 mM K^+ phosphate buffer (pH 7.0). The peaks of imino protons were assigned according to the reference.² **(B)** CD spectrum of 2.5 μM LTR-III at 25°C in 50 mM K^+ phosphate buffer (pH 7.0). **(C)** MD simulation of free LTR-III. RMSD for the solvated free LTR-III from the energy-minimized structure as a function of simulation time. **(D-E)** Overlay of snapshots from top view **(D)** and side view **(E)**. The backbone is represented as ochre ribbon, and the base is represented as light yellow ring.

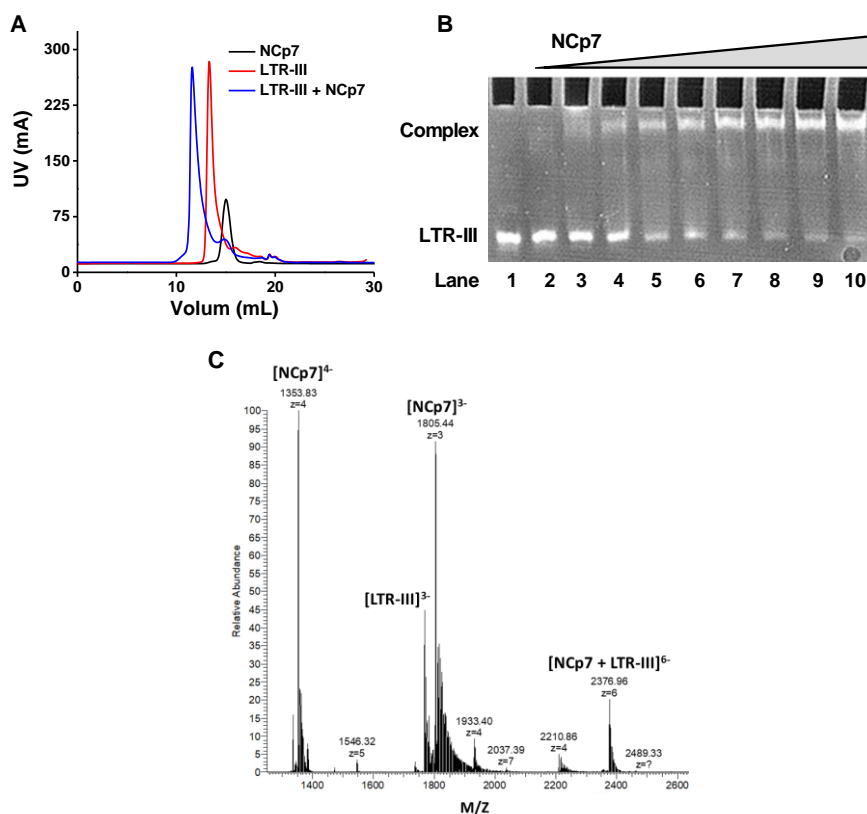


Figure S3. Interaction of NCp7 with LTR-III. **(A)** Size exclusion chromatograms of NCp7, LTR-III and the complex of NCp7/LTR-III. SEC profiles were recorded with UV detection at 280 nm. **(B)** EMSA analysis of the interaction of NCp7 with LTR-III. 5 μ M LTR-III was incubated with different concentrations of NCp7 for 30 min. Lane 1: LTR-III only; Lane 2 - 10: LTR-III + NCp7 (5, 10, 15, 20, 30, 40, 60 or 80 μ M), respectively. **(C)** ESI-MS analysis of NCp7 in the reaction with LTR-III. ESI-MS spectra were recorded to directly measure the product of LTR-III/NCp7 complex that purified by size-exclusion chromatography.

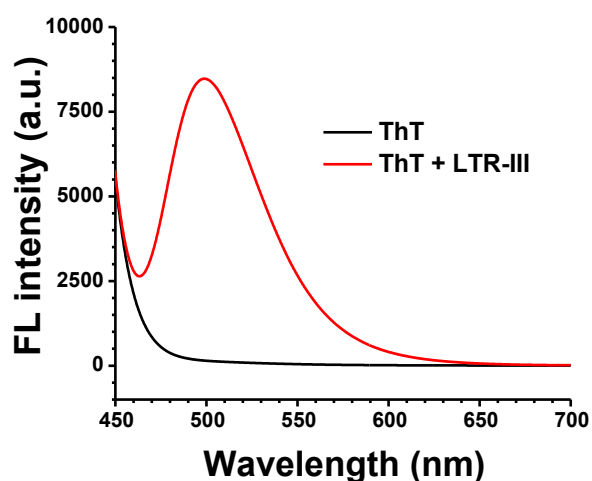


Figure S4. Fluorescence spectra of ThT in the absence (black) and presence (red) of LTR-III. Spectra were recorded on 10 μ M ThT after the incubation with 50 μ M LTR-III at 25°C for 10 min in 50 mM potassium phosphate buffer (pH 7.0).

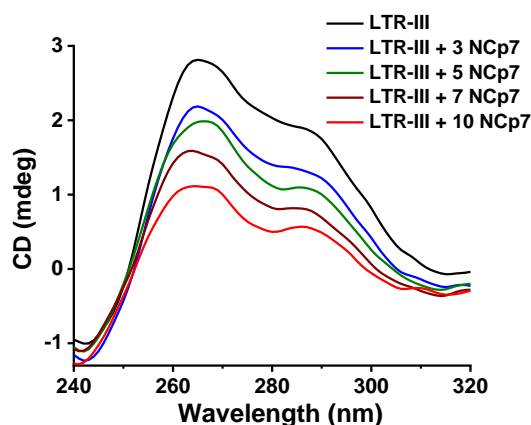


Figure S5. CD analysis of the unfolding of LTR-III upon NCp7 binding. 5 μ M LTR-III DNA G4 was incubated with different concentration of NCp7 at 37°C for 1 h (in 20 mM Tris-buffer, 140 mM KCl, pH 7.4).

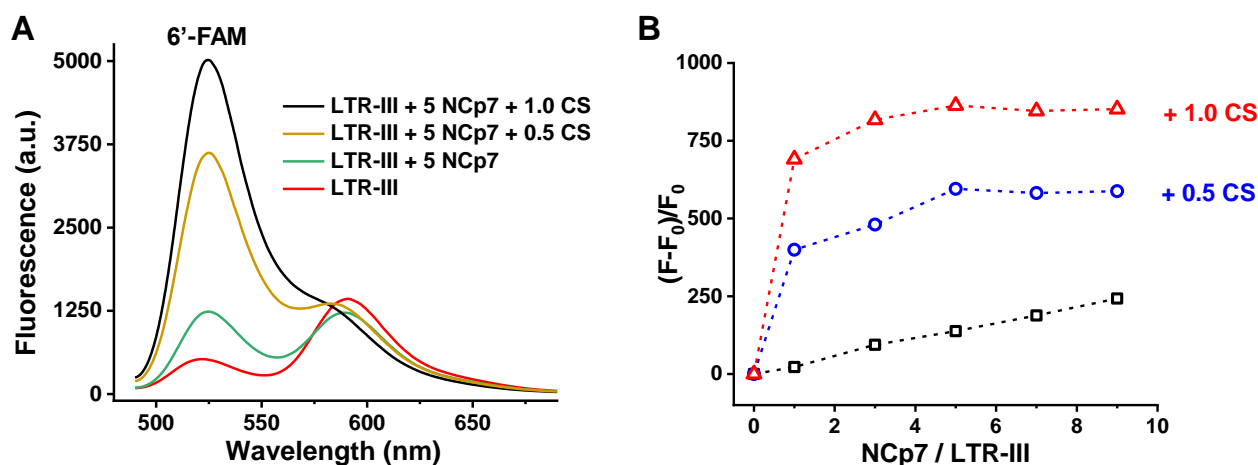


Figure S6. FRET analysis of the unfolding of F-LTR-III-T upon NCp7 binding in the presence or absence of the complementary sequence (CS). (A) The fluorescence spectra of F-LTR-III-T. The spectra were recorded on 0.6 μ M F-LTR-III-T with excitation of 6'-FAM at 480 nm. NCp7 (5 molar equivalent) and CS (0.5 or 1.0 molar equivalent) were added to analyze the unfolding of the G4 structure. (B) Fluorescence alteration of F-LTR-III-T with different amount of NCp7 in the presence or absence of CS. The relative intensity of the fluorescence (at 525 nm) was calculated using the formula $(F - F_0)/F_0$, where F_0 is the initial fluorescence of F-LTR-III-T, F is the fluorescence during the titration. The spectra were recorded in 20 mM Tris-buffer containing 140 mM KCl (pH 7.4) after 1 h incubation at 25°C.

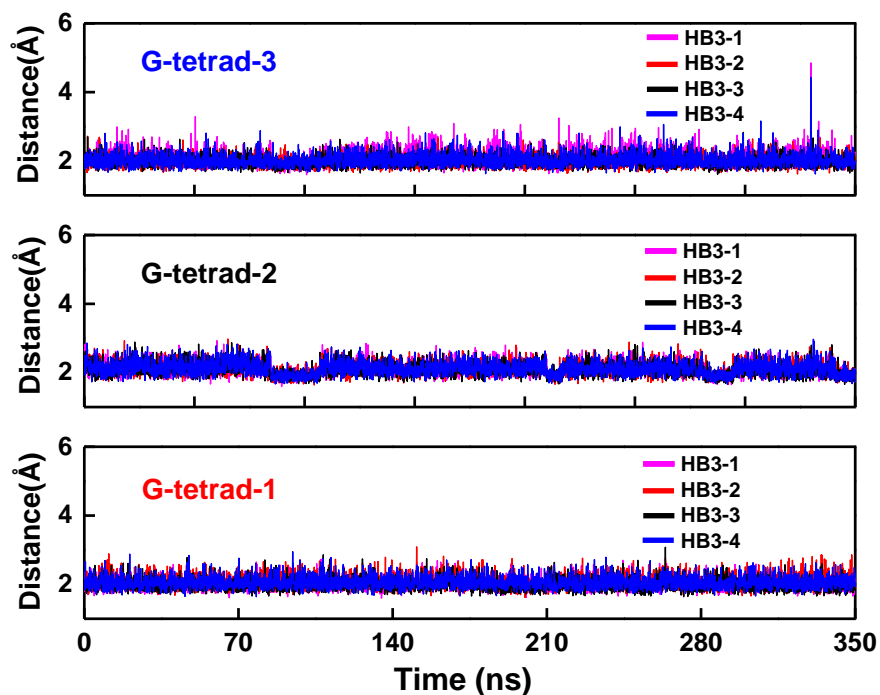


Figure S7. Hoogsteen hydrogen bond distance plots of G-tetrad-1, G-tetrad-2, and G-tetrad-3 in free LTR-III, respectively.

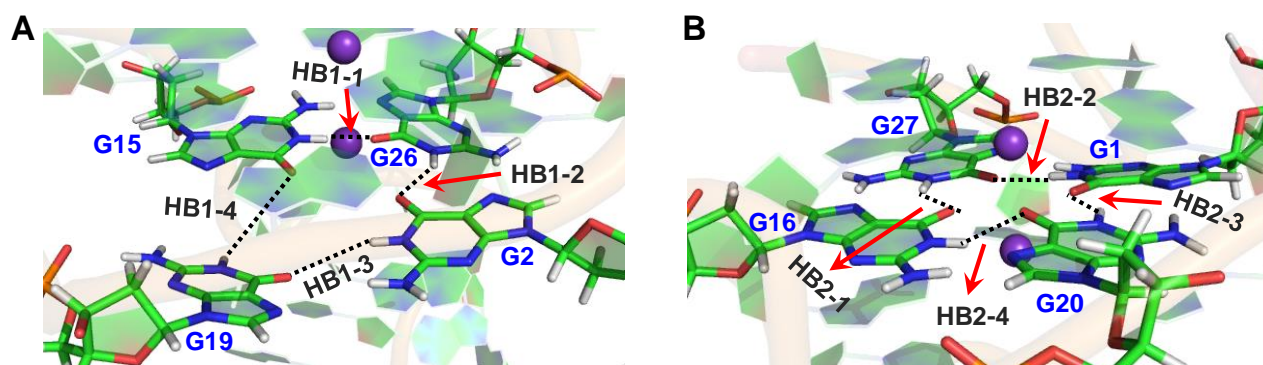


Figure S8. Hoogsteen hydrogen bond of G-tetrad of LTR-III upon NCp7 binding. (A-B) LTR-III in cartoon representation with G-tetrads-1 (A) and G-tetrads-2 (B) highlighted in stick mode, the Hoogsteen hydrogen bonds and associated bases are labeled out layer by layer.

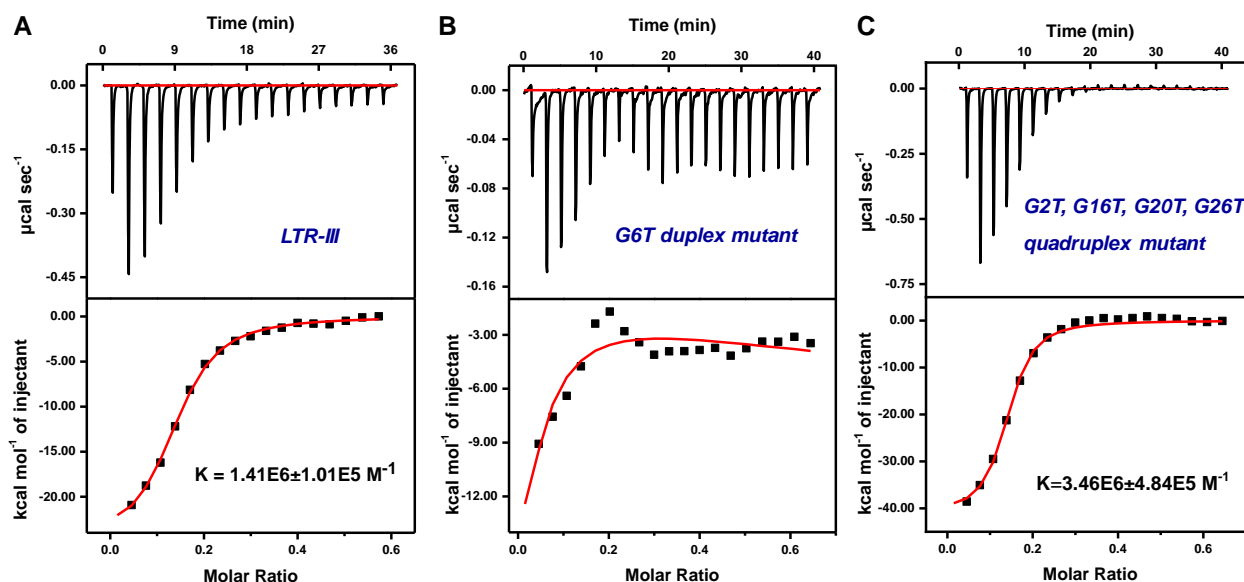


Figure S9. ITC binding isotherm of LTR-III variants to NCp7 at 25°C. (A) WT LTR-III; (B) duplex mutant (G6T); (C) quadruplex mutants (G2T, G16T, G20T, G26T). Raw heat data (top) and the integrated heat data with the nonlinear regression fits (bottom) are shown.

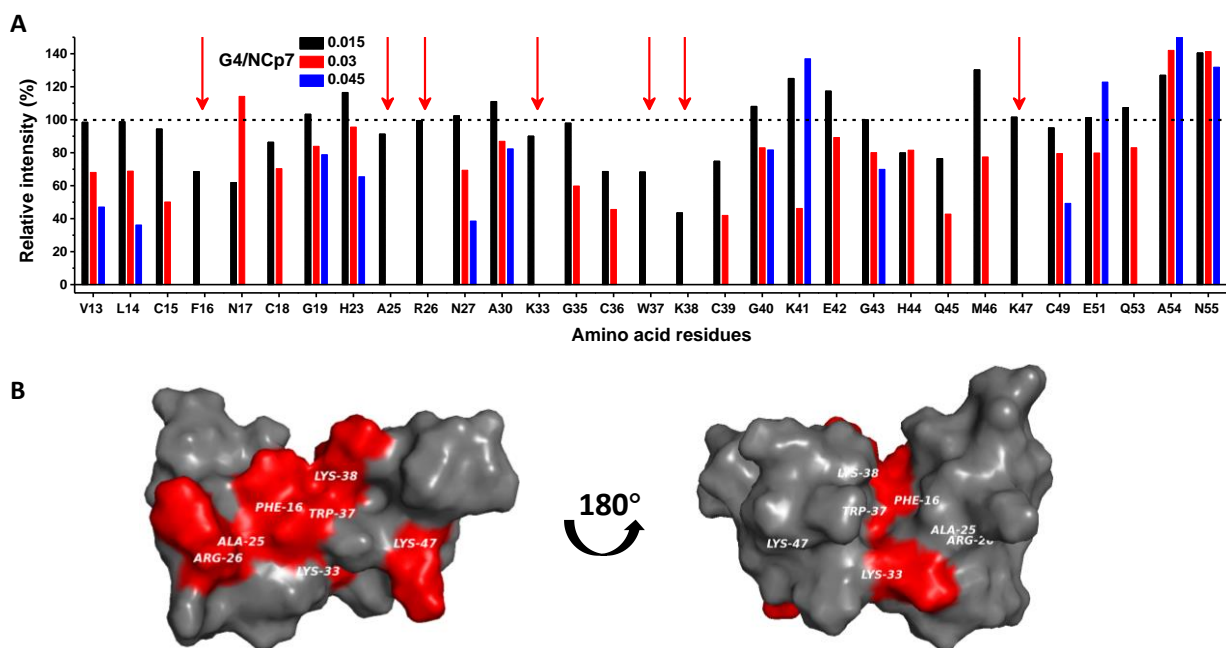


Figure S10. The region of NCp7 that involved in the interaction with LTR-III. (A) Graphical representation of intensity changes perturbation per residues of NCp7 protein in the presence of 0.015, 0.03, 0.045 eq. LTR-III. The data is driven from 2D ^1H - ^{15}N HSQC NMR titration experiments, and the residue exhibiting the largest effects has been pointed out with the arrow. (B) Residues exhibiting the largest effects in (A) have been colored in red on the structure of free NCp7 (PDB: 1ESK).

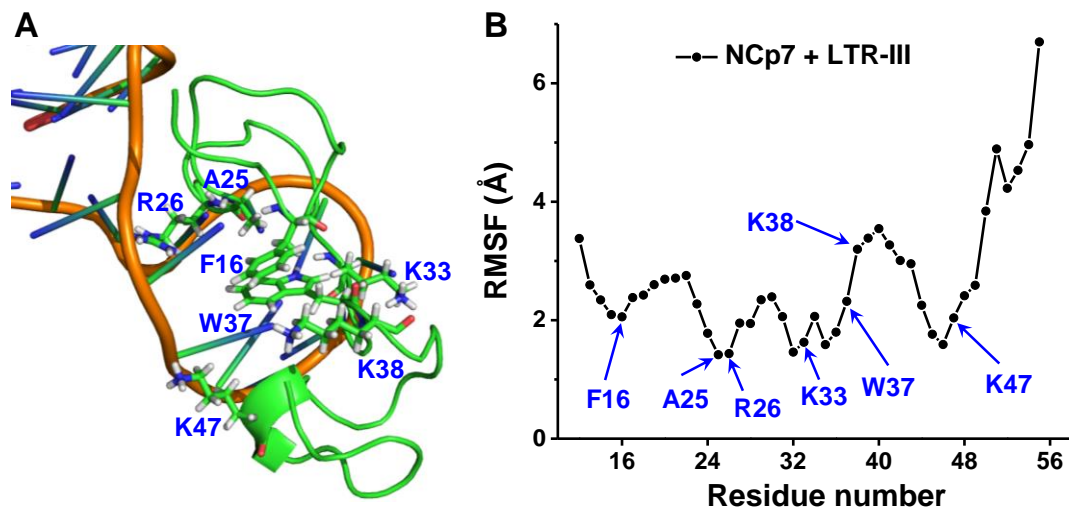


Figure S11. The possible interactions among the key residues of NCp7 that play an interaction with LTR-III. **(A)** LTR-III and NCp7 in cartoon representation with NMR signal-decreasing residues during the titration rendered as sticks and labelled out. **(B)** RMSF analysis of residues of NCp7 when interacting with LTR-III over MD simulation, with the NMR signal-decreasing residues pointed out.

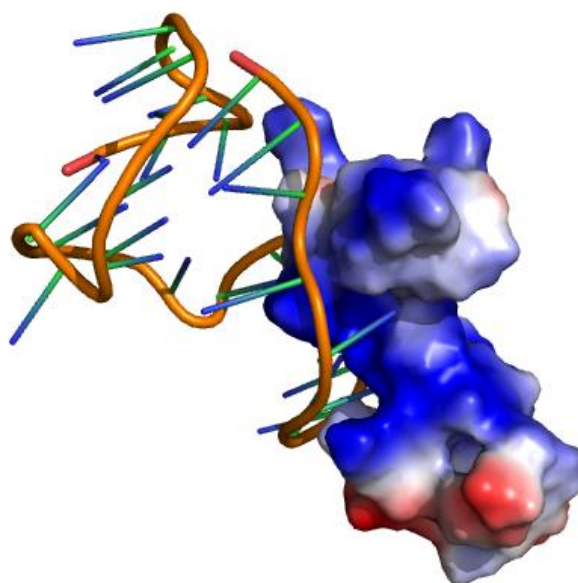


Figure S12. Binding mode of LTR-III in cartoon representation and NCp7 in electrostatic potential representation on the surface. The potential is shown from -45.1 kcal/mol/e (red) to +45.1 kcal/mol/e (blue).

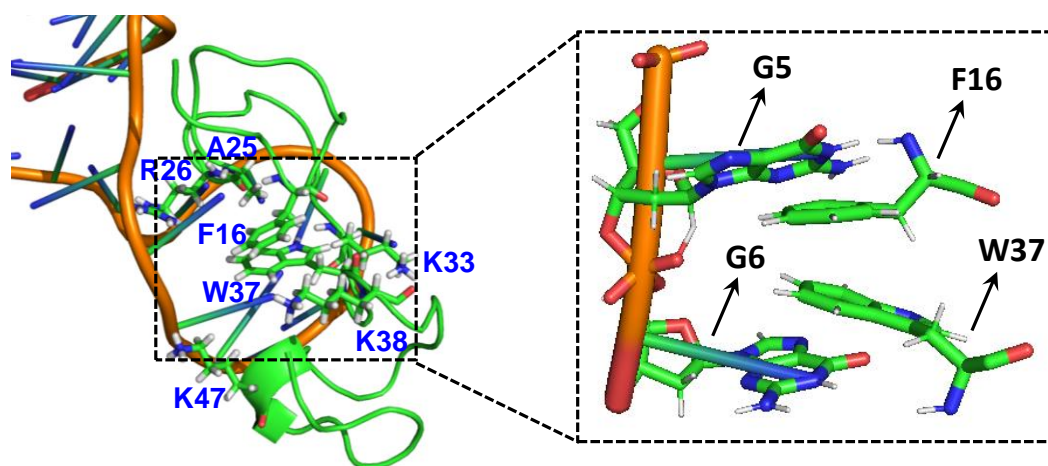


Figure S13. LTR-III and NCp7 in cartoon representation with NMR signal-decreasing residues during the titration rendered as sticks and labeled out.

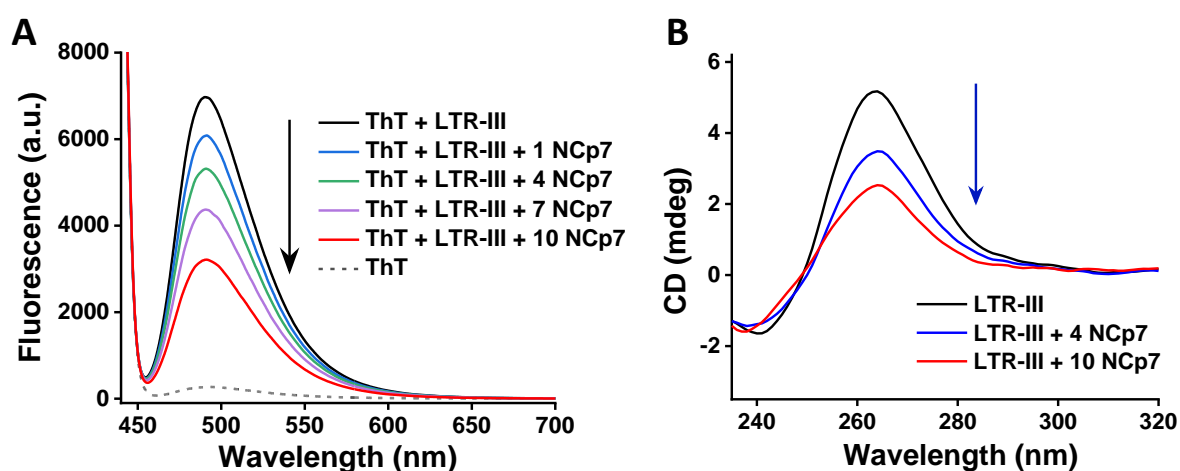


Figure S14. NCp7 unfolds the G4 structure of LTR-III RNA. (A) Fluorescence measurement of the interaction of NCp7 protein with LTR-III RNA G4 using ThT assay. NCp7 was titrated into 100 μM LTR-III containing 10 μM ThT (in 20 mM Tris-buffer, 140 mM KCl, pH 7.4). The samples were incubated for 30 min at 37°C and the spectra were recorded with the excitation of 420 nm. (B) CD analysis of the unfolding of LTR-III RNA G4 upon NCp7 binding. 5 μM LTR-III RNA G4 was incubated with different concentrations of NCp7 at 37°C for 1 h (in 20 mM Tris-buffer, 140 mM KCl, pH 7.4).