

Article

Significant Differences in RNA Structure Destabilization by HIV-1 Gag Δ p6 and NCp7 Proteins

Micah J. McCauley ^{1,*} , Ioulia Rouzina ² , Jasmine Li ³, Megan E. Núñez ³
and Mark C. Williams ^{1,*} 

¹ Department of Physics, Northeastern University, Boston, MA 02115, USA

² Department of Chemistry and Biochemistry, The Ohio State University, Center for Retroviral Research, and Center for RNA Biology, Columbus, OH 43210, USA; irouzina@gmail.com

³ Department of Chemistry and Program in Biochemistry, Wellesley College, Wellesley, MA 02481, USA; jli12@wellesley.edu (J.L.); mnunez@wellesley.edu (M.E.N.)

* Correspondence: m.mccauley@northeastern.edu (M.J.M.); ma.williams@northeastern.edu (M.C.W.); Tel.: +1-617-373-2345 (M.J.M.); +1-617-373-5705 (M.C.W.)

Received: 10 March 2020; Accepted: 21 April 2020; Published: 25 April 2020



Abstract: Retroviral nucleocapsid (NC) proteins are nucleic acid chaperones that play distinct roles in the viral life cycle. During reverse transcription, HIV-1 NC facilitates the rearrangement of nucleic acid secondary structures, allowing the transactivation response (TAR) RNA hairpin to be transiently destabilized and annealed to a complementary RNA hairpin. In contrast, during viral assembly, NC, as a domain of the group-specific antigen (Gag) polyprotein, binds the genomic RNA and facilitates packaging into new virions. It is not clear how the same protein, alone or as part of Gag, performs such different RNA binding functions in the viral life cycle. By combining single-molecule optical tweezers measurements with a quantitative mfold-based model, we characterize the equilibrium stability and unfolding barrier for TAR RNA. Comparing measured results with a model of discrete protein binding allows us to localize affected binding sites, in addition to quantifying hairpin stability. We find that, while both NCp7 and Gag Δ p6 destabilize the TAR hairpin, Gag Δ p6 binding is localized to two sites in the stem, while NCp7 targets sites near the top loop. Unlike Gag Δ p6, NCp7 destabilizes this loop, shifting the location of the reaction barrier toward the folded state and increasing the natural rate of hairpin opening by $\sim 10^4$. Thus, our results explain why Gag cleavage and NC release is an essential prerequisite for reverse transcription within the virion.

Keywords: HIV-1; TAR hairpin; Gag; nucleocapsid; optical tweezers; energy landscape; transition state; mfold

1. Introduction

DNA and RNA hairpins participate in many critical rate-limiting biochemical processes [1–5]. These sequences are often chaperoned by ligands that facilitate the formation of a diverse set of secondary and tertiary structures to suit their roles [2–8]. The transactivation response (TAR) sequence in the HIV-1 virion forms a hairpin, which is involved in several key steps in the virion life cycle (Figure 1A), including transcription and reverse transcription [9,10]. This 59-base hairpin has an unusually irregular structure (Figure 1B) where the base pairing is interrupted by several bulges and mismatches. Despite the ‘patterned’ structure of the folded hairpin, it is stable enough to facilitate the faithful transcription of the remaining genome [9]. Once the virion infects a new host, however, the hairpin must open to facilitate the key step of minus strand transfer (Figure 1C) [11–16]. TAR unfolding in this rate-limiting step must proceed without the assistance of any ATP-driven motor.

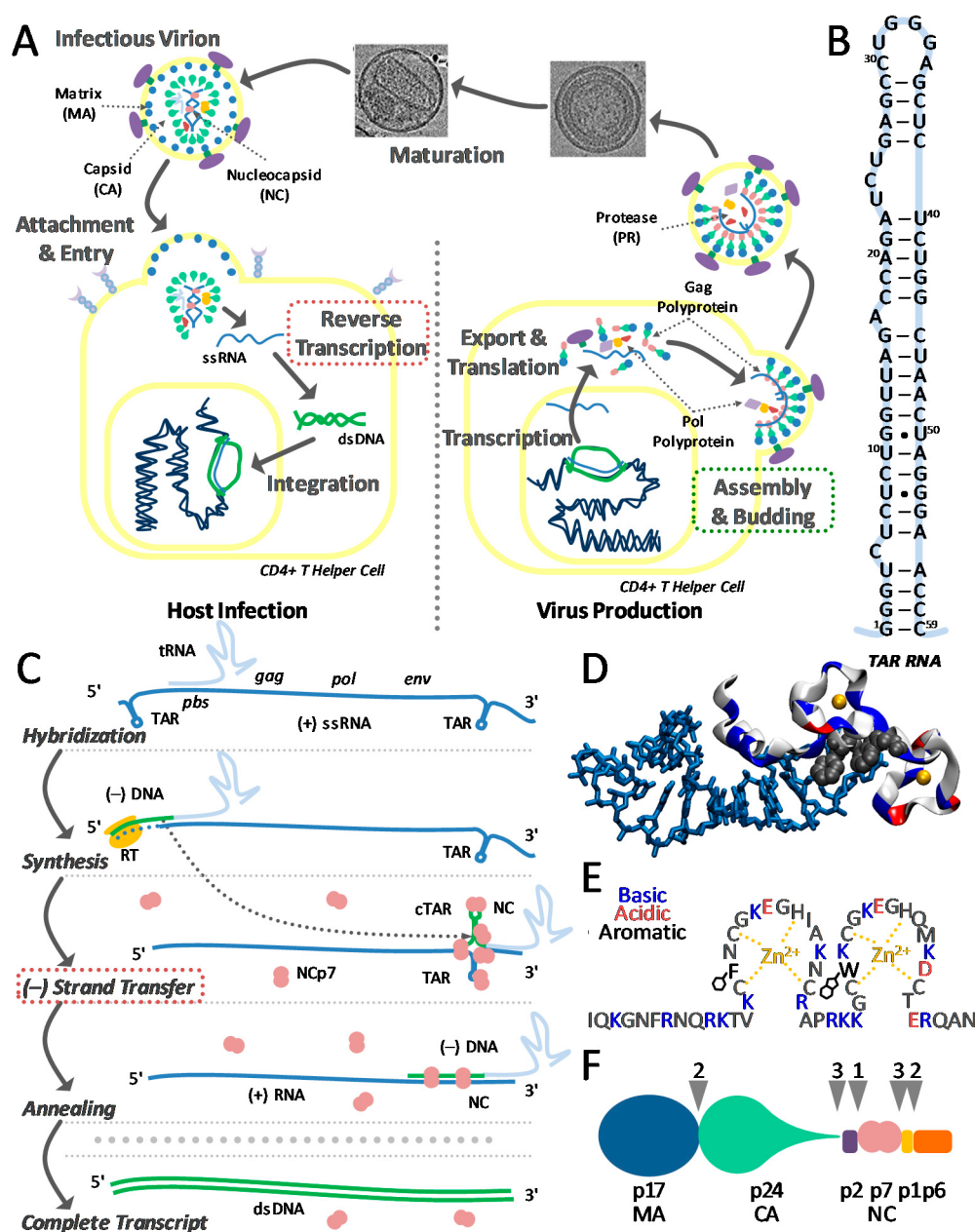


Figure 1. Critical steps in the viral life cycle. (A) In a standard picture of the infection, the virion attaches to a host cell and releases the viral RNA, which must undergo reverse transcription before integration. An infected cell will transcribe viral polypeptides, including the group-specific antigen (Gag). Polymerizing upon the host cell wall, Gag packages two RNA strands into the nascent virus. Reverse transcription may not occur until the nucleocapsid domain NCp7 is cleaved from Gag. (B) The folded structure of the 59-base transactivation response (TAR) RNA hairpin is highly irregular, with several bulges and mismatches. (C) Nucleic acid chaperone NCp7 facilitates minus strand transfer, a rate-limiting step in reverse transcription, where viral RNA (blue line) is converted into DNA (green line). (D) The sequence of NCp7, highlighting two zinc finger domains, responsible for nucleic acid binding and destabilization, and a basic tail (blue residues) responsible for chaperone activity. (E) NMR structure of the nucleocapsid (NC) bound to SL3, where two aromatic residues (black) stack with exposed bases (PDB code 1a1t) [17]. (F) Progressive cleavage (arrows) of Gag polyprotein by viral protease-free NCp7 as well as the matrix (MA) and capsid (CA) domains [18,19].

The nucleocapsid protein NCp7 consists of 55 amino acids, forming two functional subdomains (Figure 1D). A basic tail serves as an effective aggregator of DNA and RNA. The other subdomain coordinates two zinc ions, often referred to as zinc ‘fingers’, which contribute to nucleic acid binding and destabilization. Each ‘finger’ also contains an aromatic residue capable of stacking with exposed residues (this is shown for the SL3 hairpin in HIV-1 in Figure 1E) [17]. Together, these subdomains combine to form a highly effective nucleic acid chaperone, which enables the rearrangement of nucleic acid structures into an optimal form [20–22]. In a previous study, we showed that NCp7 effectively destabilizes TAR and facilitates hairpin opening by a factor of 10^4 [23]. However, NCp7 does not significantly affect the final annealed DNA–RNA hybrid [7,11].

Within an infected host cell, NCp7 is expressed as a subdomain of the group-specific antigen (Gag) protein. This polyprotein comprises four major domains: a matrix protein, a capsid protein, the nucleocapsid protein NCp7, and the viral budding p6 domain. (Figure 1F) [18]. Gag polymerizes on the host cell plasma membrane to form an immature virion, while packaging the viral RNA (Figure 1A). Between 1500–3000 Gag molecules are required to assemble a single virion [24–26]. Once the virion buds from the host, viral protease, contained as a subdomain within the Gag-Pol polyprotein (which constitutes about 5% of all Gag molecules), cleaves itself and subsequently NCp7 from the polyprotein [18]. Full cleavage of the polyproteins marks the conversion of the virus from immature to a mature, infectious virion (Figure 1F). Both NCp7 and Gag are effective at annealing tRNA₃^{Lys} to the primer-binding site on the HIV-1 genome [27], which is the first step in reverse transcription that requires nucleic acid chaperone activity [10]. However, only processed nucleocapsid (NC) facilitates minus-strand transfer [28,29]—the next step in reverse transcription that requires effective nucleic acid chaperone activity [30,31]. Why is NCp7 a more effective chaperone than its parent protein, Gag, for minus strand transfer, despite NC being the primary RNA-interacting domain for both proteins?

In this work, we use single molecule force unfolding data to measure the free energy landscape of TAR hairpin unfolding. We then quantify changes to this landscape in the presence of the GagΔp6 polyprotein. Energy landscapes are also modeled using data from mfold for the TAR sequence. Quantitative comparisons between the model and experiment allow us to deduce specific Gag binding sites along the TAR hairpin that alter the free energy of hairpin unzipping. We find that, while binding does destabilize the folded hairpin, it does not affect the lower stem, so that the transition barrier to unfolding is nearly unchanged and the natural rate of hairpin unfolding is increased by only $\sim 10\times$. This result is in sharp contrast to the effect of NCp7: once cleaved from Gag, NCp7 binds to the TAR hairpin, where it not only destabilizes the folded state, but also uniquely shifts the location of the barrier to initiate unfolding. That shift means only a short stem needs to be unfolded, which leads to a decisive increase of $\sim 10^4\times$ in the rate of unfolding [23]. Gag cleavage by the protease is thus a critical step that enables reverse transcription within the virion.

2. Materials and Methods

2.1. Hairpin Construct Synthesis

The creation of end labeled RNA hairpin/DNA handle hybrids has been outlined before [23,32]. Essentially, a 5′-biotinylated 3400 base pair DNA handle and a 3′-dioxxygenin 3100 base pair DNA handle were each created from custom primers in PCR-amplified plasmid pBR322. The 5′-biotinylated and 3′-dioxxygenin handles were digested with EcoRI and BSPEI, respectively, and then each was purified on a 0.8% agarose gel. A PCR-amplified 103-base synthetic DNA template was prepared to include the 59-base TAR sequence, the T7 promoter site and two flanking sequences. After transcription by T7 RNA polymerase, the resulting RNA was purified on a denaturing 10% PAGE gel. The RNA hairpin was ligated to short DNA oligos, using T4 RNA ligase 1 on the 5′ side and T4 RNA ligase 2 on the 3′ side (this required three ribonucleotides on the 3′ side of the oligo). Each ligation was followed by purification on a denaturing 10% PAGE gel. This construct was ligated to each handle

with T4 DNA ligase, with a 0.8% agarose gel after each ligation. Finally, these TAR hairpin constructs were then incubated with a solution of streptavidin-coated beads (3.1 microns in diameter, Spherotech, Lake Forest, IL, USA), to achieve an average ratio of six constructs per bead.

2.2. Gag Polyprotein Preparation

Gag polyproteins studied here specifically lack myristic acid at the N-terminus and the p6 domain at the C-terminus and are also commonly referred to as Gag Δ p6-Myr (though, throughout this work, they will be termed Gag Δ p6) [33–35]. Gag Δ p6 was prepared as previously described [33–35].

2.3. Single Molecule Optical Tweezers

Several papers have detailed the optical tweezers setup and have specified the utility of this design for quantifying hairpin unfolding [23,36–41]. Here, a custom-built flow cell was placed at the center of a dual 830 nm beam (Lumics, Berlin, Germany), counter propagating optical trap. Anti-dig-coated beads (2.1 microns in diameter, Spherotech) flowed into the cell, were caught in the trap and pulled onto a pre-pulled micropipette tip (0.5-micron diameter, WPI). Beads pre-incubated with the TAR hairpin (~6 hairpins for each bead) were diluted (~fM), then introduced into and caught in the trap. The experimental buffer included 10 mM HEPES, pH 7.5 and 100 mM Na⁺ (no Mg²⁺ was used), while the flow cell temperature was 22 °C. These experimental conditions are identical to experiments done on NCp7. Single tethers were isolated between the two beads and tested by force extension in the schematic shown in Figure 2A. Cycles of force extension and release were observed with a step resolution of 2 nm, at a corresponding loading rate of 10 pN/s. Figure 2B shows discrete hairpin opening events, which were characterized by opening and closing forces (F_{op} and F_{cl}) and an opening length (Δx_{op}). The work performed by the instrument during unfolding was also determined as $W_{op} = \Delta G_{ds} - \Delta G_{ds+ss}$ (the shaded region in Figure 2B), which corrects for the elastic stretching contribution of both the hairpin and the DNA handles. Bead + TAR hairpins were also exposed to 100 nM and 10 nM solutions of Gag (in the same experimental buffer, which was also the same buffer used in NCp7 experiments). Hairpin opening statistics were not measurably different in these concentrations, while the DNA handles were somewhat less affected in the lower concentration. This is consistent with previous Gag–nucleic acid binding experiments that suggest that TAR RNA is saturated with the protein [35]. Because the incubated bead + TAR RNA hairpin solution is very dilute (~fM) compared to Gag Δ p6 (10 nM), it is not the protein:RNA ratio, but rather the protein–RNA binding affinity determined by the solution’s ionic strength that determines the complex saturation that is required for protein function.

2.4. Hairpin Energies from Mfold

Theoretical landscapes were deduced from the base pair and stacking data set available at the mfold server [42,43]. Specifically, the RNA folding server (at <http://unafold.rna.albany.edu>), version 3.6, is given the 59-base TAR sequence (the UNAFold server also returns the same values under the appropriate conditions). Though the input conditions were fixed at 37 °C and 1 M Na⁺, we have previously shown that this is effectively equivalent to the experimental conditions of 22 °C and 100 mM Na⁺ in this study [23]. For the case of TAR RNA, only one hairpin structure corresponded to the minimum folded structure energy, which incorporated the full $\Delta x_f = 59$ bases (no unfolded flanking tails were left) with a total energy of $\Delta G_f = 32.9$ kcal/mol, or 55.5 k_BT (this structure may be seen in Figure 1B). Furthermore, mfold provided a sequence of energies corresponding to each paired base (N_i), including pairing and stacking contributions determined from an array of calorimetry studies; $\Delta G_i(N_i, F=0)$. These individual elements are summed to give the total hairpin energy, $\sum_i \Delta G_i(N_i, F=0) = G_N(N, F=0) = \Delta G_f$, which we used to construct the energy landscapes shown below.

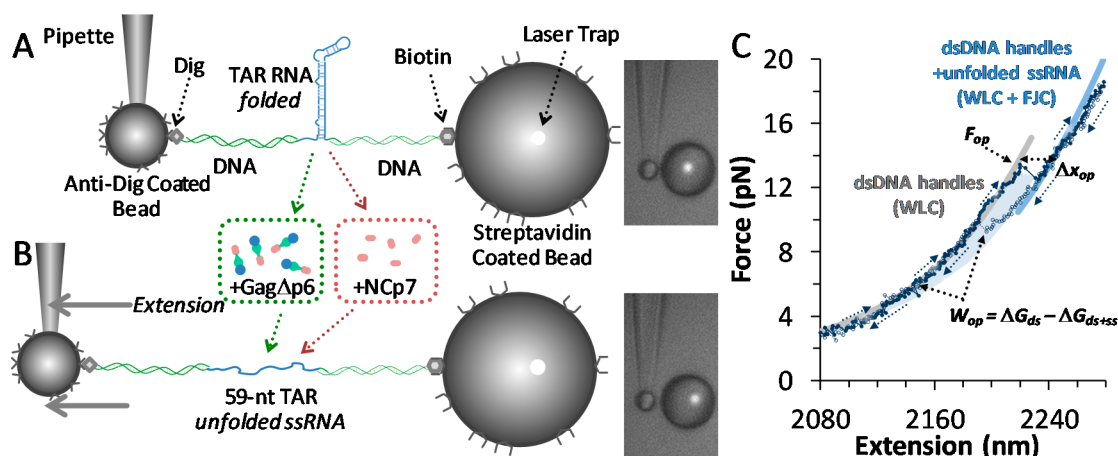


Figure 2. Single molecule assay observes hairpin opening and folding. (A) Single molecule construct of a TAR RNA hairpin (blue) flanked by long, labeled dsDNA handles (green), suspended between functionalized beads. The anti-dig-coated bead (2.1 μm diameter) was fixed onto a glass micropipette, while the streptavidin-coated bead (3.1 and 5.4 μm diameter beads were used) was held in a dual-beam optical trap. (B) Stepwise increases in extension raise the tension until a critical force is reached and the hairpin is completely disrupted, leaving ssRNA. (C) Three cycles of increasing force extension (along the arrows, filled circles) and force release (open circles decreasing along the arrows) reveal this length change (Δx_{op}), the force (F_{op}) and the work done by the instrument in hairpin unfolding (W_{op}). Models of polymer elasticity are shown and are described in Appendix A. Experimental buffers included 10 mM HEPES (pH 7.5) and 100 mM Na^+ , unless noted. Results were compared to experiments performed in saturating conditions of Gag Δ p6 and NCp7.

3. Results

Parameters extracted from optical tweezer experiments are analyzed to create a landscape of TAR hairpin unfolding. Simulated landscapes from mfold are directly compared to the data to confirm the validity of this technique. The landscapes obtained in the presence of Gag Δ p6 are compared to models of destabilization to determine the location and specificity of protein binding. Finally, Gag Δ p6 binding is directly compared to the ability of NCp7 to dramatically enhance the opening rate of the TAR hairpin, which was shown in our previous work [23].

3.1. TAR Hairpin Length Is Unaffected by Gag Δ p6

Force extension curves show a clear ‘rip’ due to the hairpin unfolding. The length of the unfolded hairpin is measured between the onset of the opening event (F_{op} , found on Figure 2C), and the same force on the unfolded construct. Subtracting the extension difference between the same forces on these curves automatically eliminates changes in the elasticity of the handles and hairpin at different forces from the result. This result for the hairpin opening length (Δx_{op}) is still a force-dependent length, and correcting for polymer elasticity yields a force-independent length change in the bases, plotted in Figure 3A. Appendix A details these models, which includes the Worm-Like Chain (WLC) used to describe dsDNA, while the Freely Jointed Chain (FJC) characterizes ssRNA. For the TAR hairpin, the unfolding length is measured to be 48.2 ± 0.6 bases. This measured length is shorter than the 59-base TAR sequence, and the discrepancy is due to the unusual amount of hairpin fraying, as explained below. Furthermore, this length did not significantly change in the presence of either Gag Δ p6 or NCp7: 47.5 ± 0.7 bases and 48.4 ± 0.5 bases, respectively (shown in Figure 3B). Thus, hairpin unfolding represents the same two-step opening length in all cases. The opening length is summarized with the rest of the measured data and is shown below in Table 1.

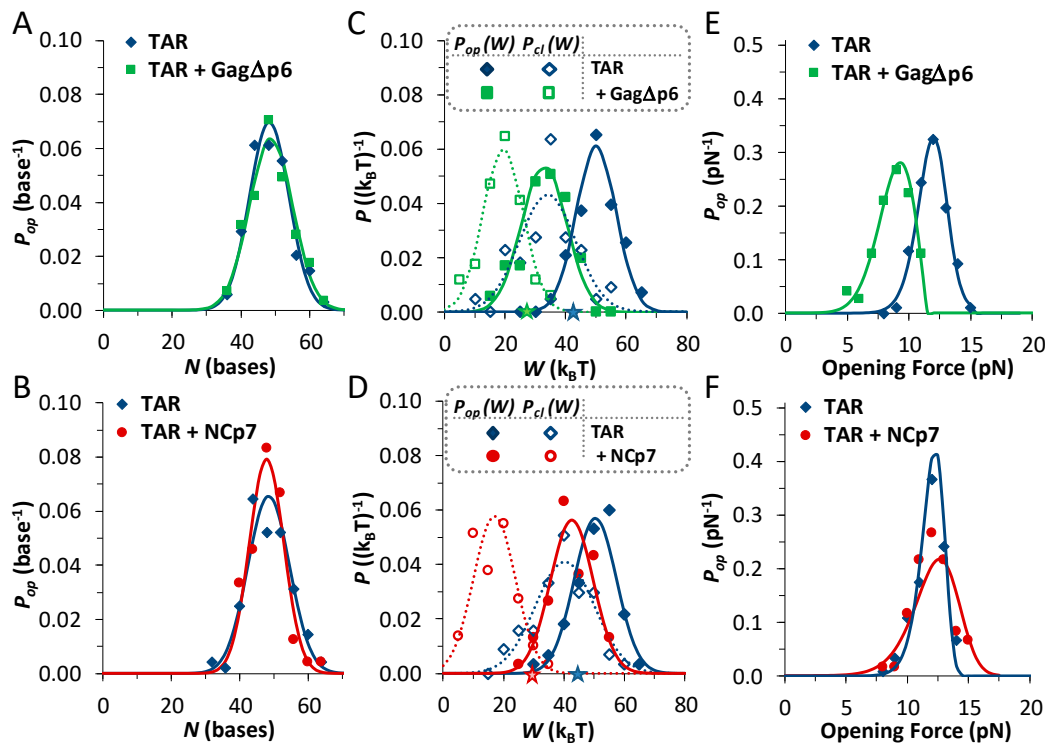


Figure 3. Force-extension/release data quantifies the energy and the barrier of unfolding. (A) The measured unfolding length (Δx_{op}) for the TAR hairpin (blue, $n = 86$ events) and for TAR in the presence of 10 nM GagΔp6 (green, $n = 72$). (B) The length for TAR alone (blue, $n = 120$) and in the presence of 50 nM NCp7 (red, $n = 60$), with no measurable change for either protein from the native hairpin. (C,D) The probability distributions of the work of unfolding/folding (solid/open symbols) cross at the energy of the hairpin unfolding (ΔG_o , marked with stars placed along the axis). Both (C) GagΔp6 and (D) NCp7 destabilize the TAR hairpin. (E,F) Distributions of the opening forces fit to the transition state model described in the text (solid lines) determining the distance to the opening barrier (Δx_{op}^+) and the barrier height (ΔG_{op}^+). Data in panels (B,D,F) were measured under the same conditions (as part of a larger data set) as in a previous work [23]. Quantified results are summarized in Table 1.

Table 1. Summary of experimental OT results and modeled mfold data for TAR unfolding and TAR in the presence of GagΔp6. Data is compared to previous results for TAR in the presence of NCp7. The opening length (Δx_{op}) and energy (ΔG_{op}) are calculated from OT and mfold data and are independent of force. The distance to the barrier (Δx_{op}^+) and the height (ΔG_{op}^+) are measured in OT experiments and calculated in mfold at the force $F_{1/2}$, where the probability of observing unfolding/folding is equal ($P_{op} = P_{cl}$). The natural rate of opening (k_{op}^0) in the absence of force was also found from OT data and reflects the instability of the TAR hairpin in the presence of NCp7 relative to the TAR and TAR + GagΔp6.

Experiment	Δx_{op} (bases)	ΔG_{op} ($k_B T$)	$F_{1/2}$ (pN)	Δx_{op}^+ (nm)	ΔG_{op}^+ ($k_B T$)	k_{op}^0 (s^{-1})
TAR (OT)	48.2 ± 0.6	42.1 ± 1.5	10.6 ± 0.2	10.9 ± 0.7	28.0 ± 1.4	$(0.7 \pm 0.4) \times 10^{-8}$
TAR (mfold)	48 ± 1	41.5 ± 1.5	10.1 ± 0.1	11.3 ± 0.9	30.7 ± 2.0	-
TAR + GagΔp6 (OT)	47.5 ± 0.7	28.9 ± 1.5	7.7 ± 0.2	9.2 ± 0.7	18.1 ± 0.9	$(10 \pm 6) \times 10^{-8}$
TAR + GagΔp6 (mfold)	48 ± 1	26.0 ± 1.5	8.0 ± 0.2	8.6 ± 1.0	20.1 ± 1.5	-
TAR + NCp7 (OT)	48.4 ± 0.5	28.3 ± 0.9	7.7 ± 0.2	4.8 ± 0.6	14.3 ± 1.3	$(1.2 \pm 0.8) \times 10^{-4}$
TAR + NCp7 (mfold)	48 ± 1	25.6 ± 1.5	7.9 ± 0.1	4.7 ± 0.9	14.8 ± 1.2	-

Results were determined from data shown in Figure 3 for TAR RNA and TAR + GagΔp6, while TAR + NCp7 was found in a previous work [23] and are shown here for comparison. In the previous work, the parameters measured for the TAR RNA hairpin were also found and were all the same as the parameters found in this study within measured uncertainty (those results are not shown). All errors are the standard error of the mean (SEM).

3.2. GagΔp6 Reduces the Energy of Unfolding

During unfolding, the instrument worked to unfold the hairpin and stretch both the dsDNA handles and the newly released ssRNA of the hairpin. Finding the work required to unfold the hairpin (W_{op}) required subtracting these elasticities. The work of unfolding is the work required to extend the folded (dsDNA only) construct from zero up to the opening force, less the energy required to extend the fully unfolded construct (dsDNA and ssRNA) over the same range ($W_{op} = \Delta G_{ds} - \Delta G_{ds+ss}$). There is also a small stiffness correction, as shown in Figure 2B. This calculation is performed for each observed unfolding and folding event and normalized distributions are shown in Figure 3C. However, these distributions are the result of ultrafast force extension/release experiments (Figure 2) and are non-equilibrium; the average work of unfolding does not give the equilibrium unfolding energy of the hairpin (ΔG_{op}). Notably, the equilibrium free energy can be found from these distributions by utilizing several techniques, as discussed in Appendix B. The unfolding energy of TAR RNA is fairly low in these experimental conditions, $\Delta G_{op} = 42.1 \pm 1.5$ k_BT, reflecting the irregular structure of this hairpin. The addition of GagΔp6 destabilizes the hairpin further, to $\Delta G_{op, Gag\Delta p6} = 28.9 \pm 1.5$ k_BT. This is similar to the effect due to NCp7 seen previously, where $\Delta G_{op, NCp7} = 28.3 \pm 0.9$ k_BT (see Figure 3D for data distributions). Thus, alone, or as a part of Gag, NCp7 further destabilizes an already unstable TAR hairpin. Measured hairpin energies are summarized below in Table 1.

3.3. GagΔp6 Lowers but Does Not Shift the Energy Barrier

Measured distributions of the opening force are plotted in Figure 3E. Fits of these distributions to a kinetic model of hairpin opening produce both the distance to the transition state ($\Delta x_{op}^{\ddagger}$) and the barrier height ($\Delta G_{op}^{\ddagger}$), as fitted parameters from the solved transition state models described in Appendix C. These fits determine that adding GagΔp6 lowers the barrier by ~10 k_BT, an effect similar to that for NCp7 (Figure 3F shows this data). However, GagΔp6 does not appear to significantly shift the transition state; $\Delta x_{op}^{\ddagger} = 10.9 \pm 0.7$ nm for TAR alone and $\Delta x_{op, Gag\Delta p6}^{\ddagger} = 9.2 \pm 0.7$ nm for TAR with GagΔp6. This contrasts with the shift induced by NCp7, where $\Delta x_{op, NCp7}^{\ddagger} = 4.8 \pm 0.6$ nm. The corresponding natural (force-free) rate of hairpin unfolding is also returned as a fitted parameter. The slow rate of hairpin opening ($k_{op}^0 = (0.7 \pm 0.4) \times 10^{-8}$ s⁻¹) is only increased by ~10× for hairpins saturated with GagΔp6 ($k_{op, Gag\Delta p6}^0 = (10 \pm 6) \times 10^{-8}$ s⁻¹). However, NCp7 induced a striking shift in the rate of opening ($k_{op, NCp7}^0 = (1.2 \pm 0.8) \times 10^{-4}$ s⁻¹), with a 10⁴x increase in the opening rate. Table 1 summarizes the fitted transition state results.

3.4. Landscape Models Locate GagΔp6 Binding

Measured landscapes are compared to an mfold-driven model of base pair/stacking stability [41,44,45]. For TAR hairpins (see Figure 4A), the energy associated with each element (canonical base pair, mismatch or loop) $G_i(N_i, F=0)$ is determined. Starting from the lower stem (force unfolding begins here), the increasing sum of these energies is found to generally increase with the length of the hairpin as it is gradually destabilized. Figure 4B shows this result for TAR, and the full opening length (Δx_f) requires the full hairpin energy (ΔG_f) in the absence of force. During optical tweezers (OT) experiments, the tension applied at the stem increases altering the opening energy along the hairpin length, with $-F \Delta x$, effectively ‘tilting’ the landscape to favor the unfolded state. At a critical force ($F_{\frac{1}{2}}$), the energies of the folded and unfolded state are equal, and the probability of seeing the folded and the unfolded state are equal ($P_{op} = P_{cl}$, for $G_i(N_i, F_{\frac{1}{2}})$). At $F_{\frac{1}{2}}$, a transition barrier may be quantified and compared to the measured data. This landscape is shown in Figure 4C, and full details of this calculation are found in Appendix D.

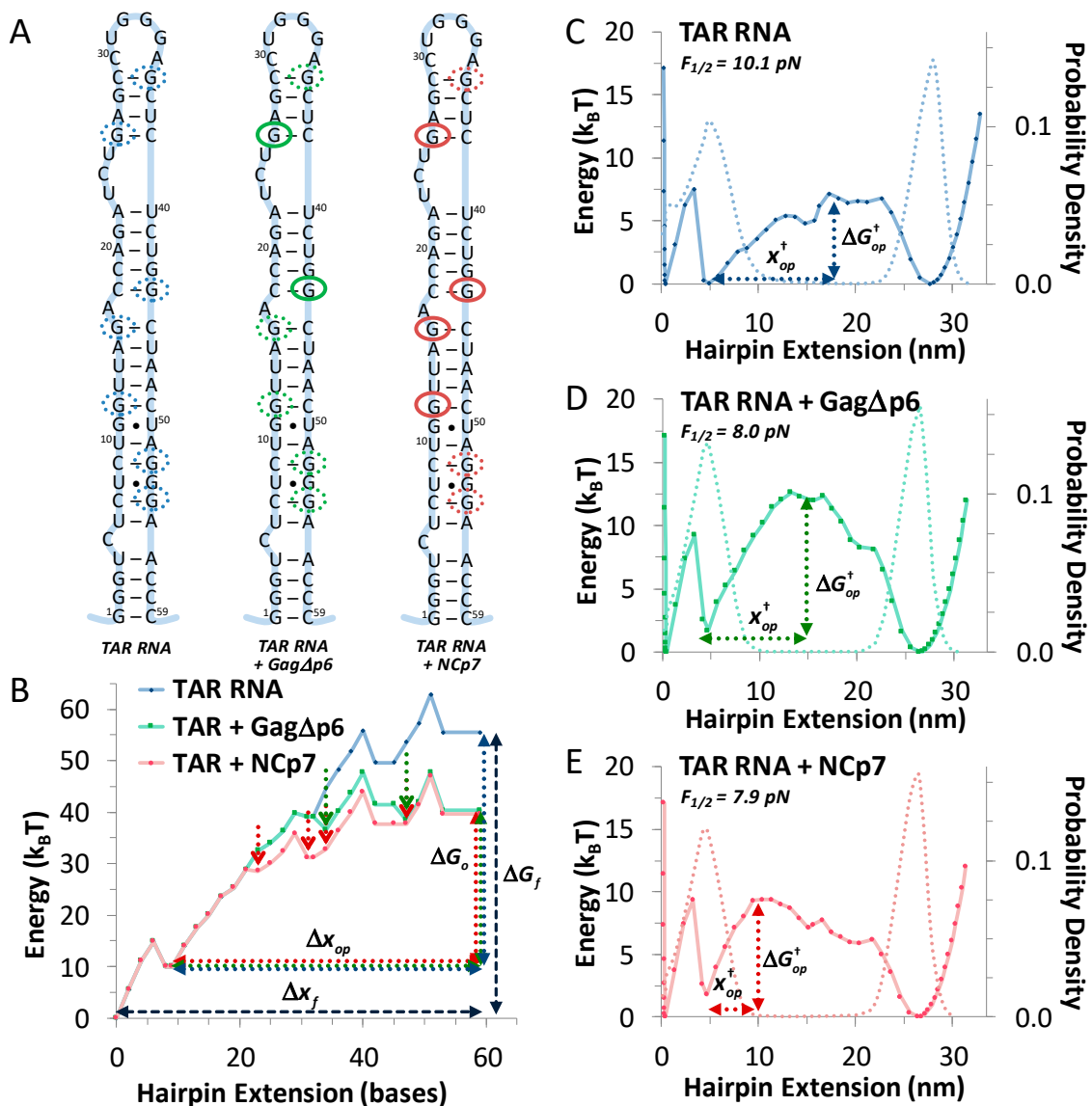


Figure 4. Quantitative modeling of critical protein binding sites. (A) Predicted mfold structure for TAR RNA hairpin, highlighting exposed G bases (dotted blue circles). GagΔp6 binding model incorporates binding to two sites (green-solid circles are confirmed sites in our model), while NC binding is shown for a four-site model (solid red circles). (B) Progressive energy of unfolding per extension, as predicted by mfold for TAR (blue), $G_i(N_i, F = 0)$. Protein binding is modeled by reducing marked sites (green and red arrows) by a discrete amount ($4 k_B T$ for NC and $8 k_B T$ at each site for GagΔp6). Predictions for the total energy (ΔG_o) and change in length (Δx_{op}) are shown, compared to the total hairpin energy (ΔG_f) and length (Δx_f) from mfold. This unfolding landscape may be 'tilted' by the effect of an external force ($-F \Delta x$). (C) Critical force landscapes for TAR (blue), (D) TAR + GagΔp6 (green) and (E) TAR + NCp7 (red), including potentials (solid lines for $G_i(N_i, F_{1/2})$, $G_{i,Gag\Delta p6}(N_i, F_{1/2})$ and $G_{i,NCp7}(N_i, F_{1/2})$) and probability densities (dotted lines), as discussed in the text. The forces, where $P_{op} = P_{cl}$ are identified for each, with values for the distance to the opening barrier ($\Delta x_{op}^{\dagger}$) and height ($\Delta G_{op}^{\dagger}$). Predicted state parameters are summarized in Table 1.

Closer examination of the landscape and the location of the unfolded/folded states (determined by the dotted probability density) shows that the lowest part of the stem is destabilized at low force. This 'fraying' happens below 4 pN (and is not directly observed in these measurements), and corresponds to the lowest 10–12 bases, which is the difference between the measured length in the experiments (Δx_{op}) and the natural length of the hairpin (Δx_f). This also corresponds to the difference

in the measured hairpin energy (ΔG_{op}) and the mfold predicted value (ΔG_f). Experimental data on an artificial DNA hairpin containing a longer stem has confirmed this effect, and there the frayed length was explicitly seen both theoretically and experimentally to unfold at a lower force [23,32]. Here, corrected for fraying, this model matches the data well; explicit comparisons are shown in Table 1 and are discussed below.

Modeling destabilization due to protein binding is somewhat more challenging. We begin with the simplified landscape data measured in OT experiments (the unfolded and the transition state) for TAR in the presence of Gag Δ p6. These data are compared to the ‘tilted landscape’ of TAR alone with the goal of altering this landscape to match the experimental data. We begin with the insight that NCp7 binds preferentially to exposed G bases along hairpin irregularities, as shown in previous binding studies [46,47]. Available sites are circled by dotted blue ovals in Figure 4A. Next, the energy difference between the measured unfolding energy for TAR and TAR in the presence of Gag Δ p6 is determined ($\Delta G_{op,Gag\Delta p6} - \Delta G_{op}$). The hairpin energy of destabilization is then applied equally to a combination over N_b binding sites ($\Delta \Delta G_{op,Gag\Delta p6}$), so that the total ($N_b \Delta \Delta G_{op,Gag\Delta p6}$) retrieves the destabilization due to Gag Δ p6 binding. For each combination, the tilted landscape ($G_{i,Gag\Delta p6}(N_i, F_{\frac{1}{2}})$) gives the position and location of the transition that may be compared to the data (see Appendix D). For Gag Δ p6, we find that destabilization is localized to just two binding sites in the upper part of the stem (Figure 4A,B,D), each destabilized by ~ 8 k_BT for the total TAR destabilization of 16 k_BT seen above. This pattern contrasts to the binding of NCp7, which, as we previously showed binds along four sites in the upper stem, completely destabilizing it (Figure 4A,B,E). There, each site was destabilized by ~ 4 k_BT for the 16 k_BT measured difference. All modeled and experimental values are compared in Table 1.

4. Discussion

4.1. Both NCp7 and Gag Δ p6 Chaperone Nucleic Acids

As described in the Methods section, the Gag Δ p6 polyprotein studied here is not the full sequence, as it lacks the both the p6 domain on the C-terminus and myristylation at the N-terminus. Thus, the C-terminus of this Gag Δ p6-Myr protein (hereafter simply called Gag Δ p6) is the NCp9 domain. Here, we compare the effect of this Gag Δ p6 construct on TAR hairpin opening to the effects of the completely processed version (NCp7), which we studied previously [23]. RNA interactions with three sequentially processed versions of HIV-1 NC: NCp15, NCp9 and NCp7 have been studied extensively. Both NCp9 and NCp7 lack the anionic unstructured C-terminal p6 domain and are very similar in their ability to bind, aggregate and chaperone the annealing of the complementary structured RNA and DNA molecules, as well as to facilitate reverse transcription [48,49]. In contrast, NCp15 (which contains p6) was shown to have a significantly compromised ability to anneal, aggregate and destabilize nucleic acid (NA) structures, as well as to support reverse transcription [48,49]. We therefore believe that the comparison between the NCp9 domain in the context of Gag and of the completely processed NCp7 protein for their ability to facilitate TAR hairpin unfolding adequately reflects the contribution of the rest of the Gag protein on the nucleic acid chaperone activity of NCp7.

Previous bulk solution studies have shown both NC and Gag strongly facilitate tRNA₃^{Lys} annealing to its primer binding site [50–52] and other complementary regions of the structured (–) strand DNA and viral RNA [31,35]. These studies agree that $\sim 100\times$ less Gag is sufficient to maximize nucleic acid annealing compared to NC. This is due to increased binding strength in physiological salt ($K_d \sim 100$ – 1000 nM for NC, and $K_d \sim 10$ nM for Gag) from additional nucleic acid interactions with the MA domain [53,54]. However, the maximal chaperone activity of Gag is still significantly weaker than NC. Furthermore, in contrast to NC, Gag does not facilitate reverse transcription [30,31]. Three major components of the chaperone activity of Gag and NC have been compared: (1) an ability to destabilize nucleic acid (NA) structures, (2) binding kinetics and (3) nucleic acid aggregating activity, which facilitates annealing [11,55]. Gag is just as good as NC in inducing aggregation and also

destabilizes nucleic acid structures [52,56]. The inefficiency of Gag compared to NC in facilitating annealing and promoting reverse transcription was associated with slower Gag–NA interaction kinetics, which was attributed to Gag–Gag interactions [30]. However, mutations and deletions in the CA domain of Gag (eliminating most of these interactions) did not allow the chaperone activity of Gag to match NC [31]. Interestingly, it was shown that addition of the polyanion inositol hexakisphosphate (IP6) to Gag–NA complexes significantly improves Gag’s chaperone function, suggesting that IP6 promotes the extended conformation of Gag, with its NC domain bound to NA and its MA domain bound to IP6, similar to the Gag conformation in the immature virion lattice [52,57]. Recent studies show that IP6 can also stabilize the extended multimerized Gag state, facilitating a six-helix bundle formation between the SP1 regions of Gag, located between its CA and NC domains [58]. This contrasts to the compact Gag conformation in which both of its cationic domains, NC and MA, are bound to the same NA molecule. It is possible that either Gag conformation or the binding of its MA domain to RNA limits the chaperone activity of Gag in its compact conformation.

4.2. While NCp7 Strongly Facilitates TAR Opening, GagΔp6 Does Not

In this work, we explicitly quantify the destabilizing effect of GagΔp6 and NCp7 on the TAR hairpin. While the process of complete TAR unzipping from its ends may not occur during (–) DNA strand transfer [8], the details of how GagΔp6 and NCp7 destabilize this long RNA hairpin with multiple structural defects is expected to contain universal information about the mechanisms by which these chaperone proteins help to restructure nucleic acids. We find that both NCp7 and GagΔp6 strongly decrease TAR RNA hairpin stability (~ 16 k_BT). The differences between GagΔp6 and NCp7 binding become evident in the transition state data. Structural irregularities in folded TAR RNA, especially those near the apical loop, cause the transition state for unfolding to occur nearly midway down the stem. In contrast, the transition state for a fully matched hairpin is located very near the top of the hairpin [32]. Nonetheless, the natural rate of TAR RNA unfolding is slow, and in the absence of any ancillary proteins, unfolding would be a rare event. The addition of GagΔp6 does destabilize the hairpin and reduce the force required to unfold TAR. However, only a few sites adjacent to the apical TAR loop are destabilized, leaving the stem end stable, the zero-force transition barrier high, and the opening rate only $\sim 10\times$ faster than in the absence of GagΔp6 (see Figure 5). In contrast, our previous results on NCp7 showed a uniform destabilization of the upper stem at multiple sites, sharply shifting the transition barrier towards the closed hairpin. Importantly, this decreased the zero-force transition barrier, and increased the natural rate of hairpin opening by $10^4\times$ (see Figure 5). This finding matched the results of bulk experiments where TAR RNA and DNA annealing increased by $10^4\times$ in the presence of NCp7 [7].

It appears that NCp7 destabilizes the TAR hairpin by binding to the majority of its available specific binding sites in the loop-adjacent apical part of TAR RNA. Here, we use the prior observation that NC, either as a separate domain or in the context of Gag, has the highest binding specificity for exposed or unstable G bases within NA stem defects. All defects are circled with blue dotted lines in Figure 5A. We can clearly see that NCp7 destabilizes TAR at all its specific sites at the top of the hairpin, thereby increasing the size of the unstable apical loop, which opens cooperatively after stem unzipping to the transition state. As in Figure 5A, in the absence of the protein chaperone, the TAR hairpin must be unzipped through about two thirds of its stem length, up to the C₁₈–G₄₄ base pair, in order for the rest of the stem to open in a single step. The addition of NCp7 leads to the destabilization of the upper part of the TAR hairpin, so that only about a third of the stem needs to be unzipped prior to the complete opening of the rest of the stem without additional force, moving the transition state to the G₁₁–U₅₀ base pair (Figure 5E). The net equilibrium unzipping free energy is equally distributed over four specific NC binding sites in the upper part of TAR beyond the transition state, with each site destabilized by ~ 4 k_BT. In contrast, the destabilizing effect due to GagΔp6 appears to be localized only to the most apical loop-adjacent specific sites, as shown in Figure 5C. TAR still must be unzipped

through about half of its stem length for the rest of the hairpin to open cooperatively, such that the transition state for unzipping now resides around the A₁₅–U₄₆ base pair.

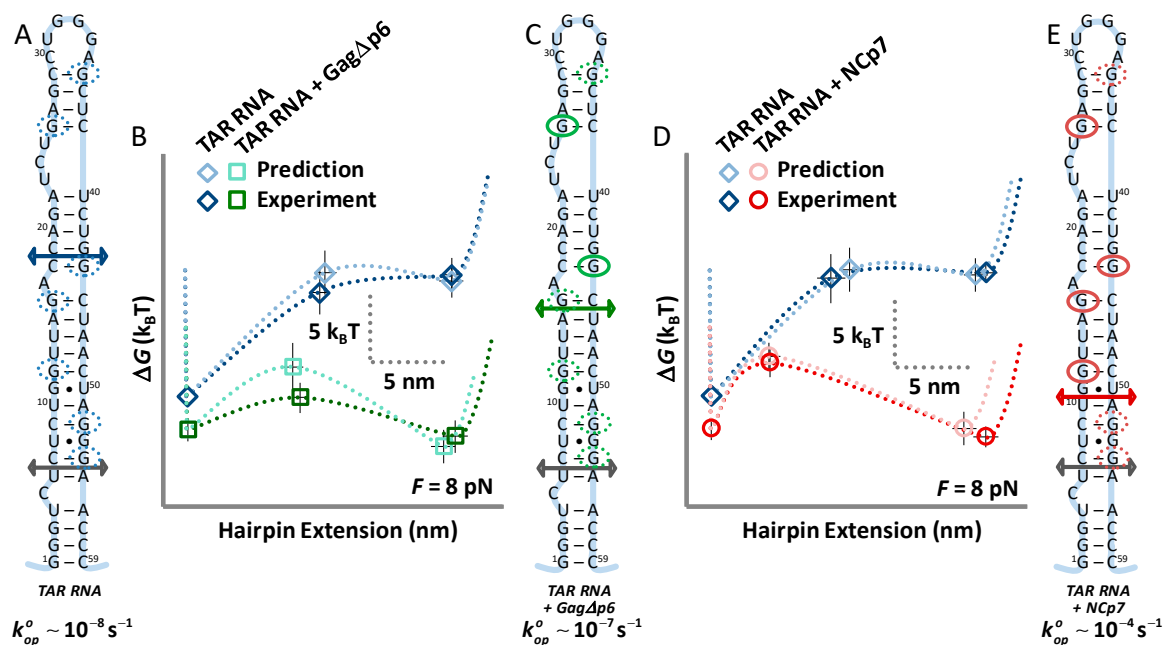


Figure 5. Unlike GagΔp6, NCp7 dramatically shifts the location of the transition barrier of the TAR hairpin. (A) The transition state of this irregular hairpin is marked by an arrow (blue), while exposed G bases are highlighted (blue dotted circles), and the natural rate of hairpin opening is low ($10^{-8} \cdot s^{-1}$). (B) The measured hairpin landscape (blue) matches values predicted from mfold (cyan), including the effects of hairpin fraying, which unfold the lowest stem (grey arrow). (C) Modeled destabilization by GagΔp6 binding is highlighted (solid green ovals show the sites destabilized by protein binding) and the resulting best match for the unfolding data is compared to TAR in (B) (data—green, model—teal). While both the energy of unfolding and the barrier decrease, the transition state is not significantly shifted (green arrow), and the opening rate is only slightly affected ($10^{-7} \cdot s^{-1}$). (D,E) NCp7 destabilizes the entire upper loop (solid red ovals are sites contributing to destabilization), leading to a marked shift in the transition state location (red arrow) toward the folded state (grey arrow) with a corresponding increase in the unfolding rate ($10^{-4} \cdot s^{-1}$, data—red, model—pink). Landscapes are shown under the effect of an 8 pN external force. TAR RNA hairpin controls measured in this work and previously are both shown to match mfold predictions.

Why would the effect of GagΔp6 on TAR stability be spatially limited, compared to NCp7, despite its lower binding specificity? There are two mutually non-exclusive possible explanations for this observation. One is that each GagΔp6 molecule is bound to TAR in its compact state with its NC and MA domains, thereby occupying a much larger binding site of approximately 20–30 nucleotides [52]. It is possible that the 59-nucleotide-long TAR RNA can only accommodate two Gag molecules, leading to the binding of only two NC domains, thereby destabilizing only the two highest affinity sites at the top of the TAR stem. The MA domain of Gag does not participate in gRNA selection for packaging through direct preferential binding and does not have chaperone activity on its own [49,59]. Therefore, only the specific binding of the NC domain of Gag, but not of MA, can lead to hairpin destabilization. At the same time, NCp7 is known to bind every ~7 nucleotides of any nucleic acid at saturation, such that multiple NCp7 molecules, up to a total of ~8–10 molecules, can bind TAR RNA. Not all these NCp7 molecules will bind specific sites and destabilize them, and the destabilization will be limited to the highest affinity locations, which, in the case of TAR, appears to be the four loop-adjacent sites. Another possible reason for the site-limited effect of GagΔp6 binding could be the higher mobility of NCp7 when bound to RNA. While detailed kinetics of NCp7 and GagΔp6

binding are unknown, it is clear that NC is highly mobile in its bound state [11]. This means that NCp7 can likely be “delocalized” over several specific sites, thereby destabilizing each of them during the timescale of our unzipping experiments (over ~2 s), while the less mobile GagΔp6 remains permanently bound on the same timescale and destabilized in only the most specific TAR locations.

Our data can be placed in the context of a recent study of Gag binding to the 5'UTR [60,61], which shows that the majority of sites within the Psi RNA region that crosslink to Gag also show significant concurrent changes in their RNA secondary structure. How is it possible that the same Gag protein binding leads to two opposite effects on the local RNA structure? The most plausible answer is that, while specific interactions between the NC domain of Gag with destabilized G residues lead to local structure destabilization, the binding of the MA domain most likely leads to a local increase in duplex stability, as it does for most non-specific multivalent cation binding interactions. Thus, the combined effect of general local destabilization and the stabilization of particular RNA bases at sites of Gag crosslinking most likely reflects the pattern of NC and MA domain binding. Previous studies of the same region of gRNA, performed on ex virio gRNA derived from infected cells in which the NC fingers in Gag were exposed to chelating agents, also shows that wild type Gag binding to RNA can either destabilize or stabilize a local RNA structure [62].

In summary, the HIV-1 NC protein, in the context of Gag, is still able to bind RNA with modest selectivity and to locally destabilize RNA structure at a few specific sites (even more strongly than NC at these sites), while leaving the rest of the RNA unperturbed, or even slightly more stable, thereby leaving high free energy barriers to NA re-folding, and leading to much longer times for NA rearrangement to achieve its most stable conformation. This contrasts with the processed NCp7 protein, which binds and slightly destabilizes the majority of its weakly specific sites, most likely by moving between these sites on a relatively fast time scale and leading to the significant facilitation of the NA restructuring rate. While the first activity of NC in the context of Gag is optimal for selective gRNA packaging, the second activity of the processed NCp7 is optimized for facilitating the refolding of any nucleic acid into its optimal lowest free energy conformation. Therefore, these results demonstrate how NCp7, only after cleavage from Gag, is optimized for nucleic acid chaperone activity during reverse transcription.

Author Contributions: J.L. and M.E.N. created TAR hairpins, while J.L. and M.J.M. performed OT experiments and M.J.M. created the mfold models. M.J.M. and I.R. analyzed the data. M.J.M., I.R. and M.C.W. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants NIH GM072462 and NSF MCB-1243883 to M.C.W. The APC was funded by Northeastern University. GagΔp6 was kindly provided by the Musier–Forsyth lab of the Ohio State University.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A Polymer Models of Elasticity

The elastic response of our long dsDNA handles under tension are modeled with the extensible Worm-Like Chain, which was solved in the high-force limit [63–65]:

$$b_{ds}(F) = B_{ds} \left[1 - \frac{1}{2} \left(\frac{k_B T}{P_{ds} F} \right)^{1/2} + \frac{F}{S_{ds}} \right]. \quad (\text{A1})$$

The measured extension under force ($b_{ds}(F)$) may be modeled to the data through fitted parameters that have been found for dsDNA over many experimental conditions, including changing salt, pH and known variations due to DNA length and trap stiffness [66–69]. The handles used in these studies are similar to those used previously and are well described by the fitted polymer parameters: $B_{ds} = 0.340 \pm 0.001$ nm/bp, $P_{ds} = 40 \pm 3$ nm and $S_{ds} = 800 \pm 200$ pN [23,32]. These values were fixed throughout the analysis, and effectively describe the extension data of the two dsDNA handles at forces below the opening transition (Figure 2B).

The Freely Jointed Chain Model is appropriate for polymers with greater rotational freedom at each element, and is commonly used to characterize single stranded nucleic acids [65]:

$$b_{ss}(F) = B_{ss} \left[\coth \left(\frac{2P_{ss}F}{k_B T} \right)^{1/2} - \frac{1}{2} \frac{k_B T}{P_{ss}F} \right] \cdot \left[1 + \frac{F}{S_{ss}} \right]. \quad (\text{A2})$$

Though there is a small range of values that adequately describe RNA, the values used here, $B_{ss} = 0.590 \pm 0.001$ nm/bp, $P_{ss} = 1.4 \pm 0.3$ nm and $S_{ss} = 800 \pm 100$ pN have been used before and give consistent results [70]. This model describes the 59-base TAR RNA. A linear combination of this with the Worm-Like Chain (WLC) above describes the extension curves after unfolding ($b_{ds}(F) + b_{ss}(F)$, Figure 2B). In practice, the total change in extension upon unfolding (Δx_{op}) is easily discerned by finding $b_{ds}(F_{op}) + b_{ss}(F_{op})$ and subtracting $b_{ds}(F_{op})$. From this force-dependent length, the number of bases (independent of force) may be recovered (Figure 3A). This approach may also apply to cycles of extension and release where the elasticity of the handles is difficult to model due to nonspecific interactions with the beads.

Comparisons with hairpins saturated with NCp7 are taken from a previous work, though that work covered several loading rates. To simplify the comparison, only data taken at the same loading rate as this work (10 pN/s) was shown from the full set. Furthermore, the data for the native TAR RNA hairpins found here matches the results from previous studies within parameters of uncertainty [23,70].

Appendix B Equilibrium Energies from Non-Equilibrium Measurements

As shown in the Methods section, the work performed by the instrument during unfolding (W_{op}) and folding (W_{cl}) is determined for each cycle. Though these measurements exclude contributions due to polymer elasticity, they are not equilibrium measurements of the hairpin energy. To recover the equilibrium value, probability density distributions were collected for work measured during folding $P_{cl}(W)$ and unfolding $P_{op}(W)$. Importantly, the integral of these distributions over the x -axis is unity. If these experiments were in equilibrium, the peak of these distributions would represent the free energy of the hairpin. Instead, these non-equilibrium distributions cross at precisely the equilibrium energy of hairpin unfolding ΔG_o [71,72]:

$$\frac{P_{op}(W)}{P_{cl}(W)} = e^{(W - \Delta G_o)/k_B T}. \quad (\text{A3})$$

These crossing points are labeled in Figure 3B along the energy axis. Though other techniques may be used to deduce the equilibrium energy, they do not appear to deliver more precise results for the data shown here [73,74]. Due to handle-bead interactions and Gag's affinity for single-stranded RNA, closing events were not always seen at the loading rates shown here. However, the hairpin energy of TAR matches that seen previously (within uncertainty) and the energy in the presence of GagΔp6 appears close to the value seen in NCp7 (the RNA binding domain of GagΔp6) [23,70].

Appendix C Kinetic Models and the Transition State

The distributions of the measured opening forces will give information about the distance to the barrier ($\Delta x_{op}^\ddagger$) and the height ($\Delta G_{op}^\ddagger$) of the barrier to unfolding [75–77]. For a simple two-state model of hairpin unfolding (as seen in this work), exact solutions to the reaction rate (in this case, the rate of hairpin opening) can be solved, even under the influence of external force ($k_{op}(F)$), which increases at a known rate (r , which is fixed at 10 pN/s in these experiments),

$$k_{op}(F) = k_{op}^o \left(1 - \frac{x_{op}^\ddagger F}{\Delta G_{op}^\ddagger} \right)^{1/\nu - 1} \exp \left\{ \frac{\Delta G_{op}^\ddagger}{k_B T} \left[1 - \left(1 - \frac{x_{op}^\ddagger F}{\Delta G_{op}^\ddagger} \right)^{1/\nu} \right] \right\}, \quad (\text{A4})$$

and the corresponding probability of observing hairpin opening at any force ($P_{op}(F)$) is then,

$$P_{op}(F) = \frac{k_{op}(F)}{r} \exp \left\{ \frac{k_{op}^0}{x_{op}^{\dagger} r} - \frac{k_{op}(F)}{x_{op}^{\dagger} r} \left(1 - \frac{x_{op}^{\dagger} F}{\Delta G_{op}^{\dagger}} \right)^{1-1/\nu} \right\}. \quad (A5)$$

The zero force (natural) rate of hairpin opening (k_{op}^0) is another independent fitting parameter, though, in practice, limited combinations of the rate, the barrier height, and the distance are physically possible. Both the rate and the probability bear explicit solutions only for two specific potential surfaces ($\nu = 1/2$ for a cubic potential and $\nu = 2/3$ for a cusp shaped potential). Furthermore, these equations are difficult to fit and non-physical local minima may be seen during minimization over a robust parameter space. To speed up the fitting process, simplified versions of these expressions relate the peak of the force distribution ($P_{op}^{\max}$, with units of pN^{-1} , and located at the force $F_{op}^{\max}$ on Figure 3E);

$$\frac{\Delta G_{op}^{\dagger}}{k_B T} = \frac{F_{op}^{\max} x_{op}^{\dagger}}{k_B T} \left[\frac{x_{op}^{\dagger}}{x_{op}^{\dagger} - P_{op}^{\max} k_B T \cdot e} (1 - \nu) \right]. \quad (A6)$$

This expression gives reasonable starting values for the barrier height ($\Delta G_{op}^{\dagger}$) and distance ($\Delta x_{op}^{\dagger}$), in the fits for the final values. Fits are shown in Figure 3C and fitted parameters are found in Table 1, while all fits yield typical values of $\chi_v^2 \sim 1$. Typically, fits to the different shape factors give different fitted results, and it is customary to fit to both factors and average the results. The averages are shown in Table 1, with propagated uncertainty. Averaged parameters for TAR RNA matched previous results within uncertainty [23]. For the shape factor $\nu = 2/3$, fits to the TAR + GagΔp6 distributions returned parameters with higher uncertainty, as there were many local minima with similar values of χ_v^2 . This may be due to GagΔp6 binding to handles and to the unfolded hairpin. Thus, a weighted average was used to combine the data from the two shape factors. Distributions of the hairpin closing forces are not characterized here, as GagΔp6 binding to ssRNA complicates a multi-step re-folding pathway [78].

Appendix D Modeling the Energy Landscape

As discussed in the text, mfold-based models derive the predicted values of the barrier location and height to compare to fitted data of the unfolding forces. Figure 4B shows the summed energies $G_i(N_i, F=0)$, and Figure 4C shows them for the TAR hairpin under the influence of force, and $G_i(N_i, F_{\frac{1}{2}})$ is where the probability of finding the hairpin open or closed is equal. The landscapes are modified to include a Morse potential $\sim D \cdot (1 - e^{-ax})^2$ and the elasticity of ssRNA $\sim k \cdot x^2$. This modified landscape is then used to calculate the probability densities of each state ($p \sim e^{-G_i(F)}$, and this is integrated over the state) [44,45,79]. This critical force landscape also highlights the lower stem fraying, as discussed in the text.

Various combinations of G-base destabilizations give changed landscapes ($G_{i,Gag\Delta p6}(N_i, F_{\frac{1}{2}})$ and $G_{i,NCp7}(N_i, F_{\frac{1}{2}})$) and different combinations of the barrier location and height ($\Delta x_{op}^{\dagger}$ and $\Delta G_{op}^{\dagger}$) that are compared to the data. The location of destabilizing sites may be quickly deduced from the single molecule data for the hairpin compared to the hairpin under the influence of the ligand. Generally, ligand binding distributed evenly along the hairpin chiefly alters the energy of unfolding (this work cannot resolve different affinities between different sites). Ligand binding localized near the lower stem induces further fraying and reduces all the measured parameters, including the observed unfolding length. However, if binding is concentrated near the apical loop, the unfolding length is not affected, while the distance to the barrier is reduced, as only the lowest part of the stem needs to be destabilized [79]. The data for TAR RNA in the presence of GagΔp6 and NCp7 reveals changes in the transition state (and unfolding energy), but no change in the unfolding length, with the difference lying in the negligible amount of transition state shift. Therefore, binding, which is already restricted to exposed G-bases, must occur at those bases near the loop. The best fits for both cases are shown in

Figure 5. In those cases, multiple combinations of sites were tried, but those shown in Figure 5 best matched the data for the transition state [23]. A four-site model with ~ 4 k_BT destabilization at each site (to match the total measured ~ 16 k_BT for NCp7 binding) best fits the data for NC. This binding model destabilizes the upper loop and reproduces the transition state shift. However, only a two-site model works best for GagΔp6, with ~ 8 k_BT destabilization at each site (to, again, match the total measured ~ 16 k_BT energy difference due to GagΔp6 binding). This model reduces the transition state height, but only modestly changes the transition state location.

References

1. Cech, T.R.; Zaug, A.J.; Grabowski, P.J. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981**, *27*, 487–496. [\[CrossRef\]](#)
2. Anthony, P.C.; Perez, C.F.; García-García, C.; Block, S.M. Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1485–1489. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Greenleaf, W.J.; Frieda, K.L.; Foster, D.A.N.; Woodside, M.T.; Block, S.M. Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science* **2008**, *319*, 630–633. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Li, P.T.; Bustamante, C.; Tinoco, I., Jr. Unusual mechanical stability of a minimal RNA kissing complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15847–15852. [\[CrossRef\]](#)
5. Onoa, B.; Dumont, S.; Liphardt, J.; Smith, S.B.; Tinoco, I., Jr.; Bustamante, C. Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme. *Science* **2003**, *299*, 1892–1895. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Glucksmann-Kuis, M.A.; Dai, X.; Markiewicz, P.; Rothman-Denes, L.B. E. coli SSB activates N4 virion RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. *Cell* **1996**, *84*, 147–154. [\[CrossRef\]](#)
7. Vo, M.N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J. Mol. Biol.* **2006**, *363*, 244–261. [\[CrossRef\]](#)
8. Vo, M.N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K. HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop “kissing” to “zipper”. *J. Mol. Biol.* **2009**, *386*, 789–801. [\[CrossRef\]](#)
9. Berkhout, B.; Jeang, K.T. 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.* **1992**, *66*, 139–149. [\[CrossRef\]](#)
10. Levin, J.G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: Critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid. Res. Mol. Biol.* **2005**, *80*, 217–286. [\[CrossRef\]](#)
11. Cruceanu, M.; Gorelick, R.J.; Musier-Forsyth, K.; Rouzina, I.; Williams, M.C. Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein’s nucleic acid chaperone function. *J. Mol. Biol.* **2006**, *363*, 867–877. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Darlix, J.-L.; Lapadat-Tapolsky, M.; De Rocquigny, H.; Roques, B.P. First Glimpses at Structure-function Relationships of the Nucleocapsid Protein of Retroviruses. *J. Mol. Biol.* **1995**, *254*, 523–537. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Lapadat-Tapolsky, M.; de Rocquigny, H.; Van Gent, D.; Roques, B.; Plasterk, R.; Darlix, J.-L. Interaction between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.* **1993**, *21*, 831–839. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Lapadat-Tapolsky, M.; Pernelle, C.; Borie, C.; Darlix, J.L. Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res.* **1995**, *23*, 2434–2441. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Thomas, J.A.; Gorelick, R.J. Nucleocapsid protein function in early infection processes. *Virus Res.* **2008**, *134*, 39–63. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Wu, H.; Mitra, M.; McCauley, M.J.; Thomas, J.A.; Rouzina, I.; Musier-Forsyth, K.; Williams, M.C.; Gorelick, R.J. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein’s nucleic acid chaperone activity and retroviral replication. *Virus Res.* **2013**, *171*, 263–277. [\[CrossRef\]](#)

17. De Guzman, R.N.; Wu, Z.R.; Stalling, C.C.; Pappalardo, L.; Borer, P.N.; Summers, M.F. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 psi-RNA Recognition Element. *Science* **1998**, *279*, 384–388. [\[CrossRef\]](#)
18. Chen, N.; Morag, A.; Almog, N.; Blumenzweig, I.; Dreazin, O.; Kotler, M. Extended nucleocapsid protein is cleaved from the Gag-Pol precursor of human immunodeficiency virus type 1. *J. Gen. Virol.* **2001**, *82*, 581–590. [\[CrossRef\]](#)
19. Shehu-Xhilaga, M.; Krausslich, H.G.; Pettit, S.; Swanstrom, R.; Lee, J.Y.; Marshall, J.A.; Crowe, S.M.; Mak, J. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J. Virol.* **2001**, *75*, 9156–9164. [\[CrossRef\]](#)
20. Rein, A.; Henderson, L.E.; Levin, J.G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: Significance for viral replication. *Trends Biochem. Sci.* **1998**, *23*, 297–301. [\[CrossRef\]](#)
21. Tsuchihashi, Z.; Brown, P.O. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1994**, *68*, 5863–5870. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Belfetmi, A.; Zargarian, L.; Tisne, C.; Sleiman, D.; Morellet, N.; Lescop, E.; Maskri, O.; Rene, B.; Mely, Y.; Fosse, P.; et al. Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. *RNA* **2016**, *22*, 506–517. [\[CrossRef\]](#)
23. McCauley, M.J.; Rouzina, I.; Manthei, K.A.; Gorelick, R.J.; Musier-Forsyth, K.; Williams, M.C. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 13555–13560. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Vogt, V.M.; Simon, M.N. Mass determination of Rous sarcoma virus virions by scanning transmission electron microscopy. *J. Virol.* **1999**, *73*, 7050–7055. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Briggs, J.A.; Simon, M.N.; Gross, I.; Krausslich, H.G.; Fuller, S.D.; Vogt, V.M.; Johnson, M.C. The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **2004**, *11*, 672–675. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Chertova, E.; Chertov, O.; Coren, L.V.; Roser, J.D.; Trubey, C.M.; Bess, J.W., Jr.; Sowder, R.C., II; Barsov, E.; Hood, B.L.; Fisher, R.J.; et al. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* **2006**, *80*, 9039–9052. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Roldan, A.; Warren, O.U.; Russell, R.S.; Liang, C.; Wainberg, M.A. A HIV-1 minimal gag protein is superior to nucleocapsid at in vitro annealing and exhibits multimerization-induced inhibition of reverse transcription. *J. Biol. Chem.* **2005**, *280*, 17488–17496. [\[CrossRef\]](#)
28. Guo, J.; Wu, T.; Anderson, J.; Kane, B.F.; Johnson, D.G.; Gorelick, R.J.; Henderson, L.E.; Levin, J.G. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* **2000**, *74*, 8980–8988. [\[CrossRef\]](#)
29. Guo, J.; Wu, T.; Kane, B.F.; Johnson, D.G.; Henderson, L.E.; Gorelick, R.J.; Levin, J.G. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **2002**, *76*, 4370–4378. [\[CrossRef\]](#)
30. Levin, J.G.; Mitra, M.; Mascarenhas, A.; Musier-Forsyth, K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* **2010**, *7*, 754–774. [\[CrossRef\]](#)
31. Wu, T.; Datta, S.A.; Mitra, M.; Gorelick, R.J.; Rein, A.; Levin, J.G. Fundamental differences between the nucleic acid chaperone activities of HIV-1 nucleocapsid protein and Gag or Gag-derived proteins: Biological implications. *Virology* **2010**, *405*, 556–567. [\[CrossRef\]](#)
32. McCauley, M.J.; Furman, L.; Dietrich, C.A.; Rouzina, I.; Nunez, M.E.; Williams, M.C. Quantifying the stability of oxidatively damaged DNA by single-molecule DNA stretching. *Nucleic Acids Res.* **2018**, *46*, 4033–4043. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Campbell, S.; Rein, A. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J. Virol.* **1999**, *73*, 2270–2279. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Campbell, S.; Vogt, V.M. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: Identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J. Virol.* **1997**, *71*, 4425–4435. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Cruceanu, M.; Urbaneja, M.A.; Hixson, C.V.; Johnson, D.G.; Datta, S.A.; Fivash, M.J.; Stephen, A.G.; Fisher, R.J.; Gorelick, R.J.; Casas-Finet, J.R.; et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.* **2006**, *34*, 593–605. [\[CrossRef\]](#)
36. Chaurasiya, K.R.; Paramanathan, T.; McCauley, M.J.; Williams, M.C. Biophysical characterization of DNA binding from single molecule force measurements. *Phys. Life Rev.* **2010**, *7*, 299–341. [\[CrossRef\]](#)

37. Greenleaf, W.J.; Woodside, M.T.; Abbondanzieri, E.A.; Block, S.M. Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.* **2005**, *95*, 208102. [[CrossRef](#)]
38. Li, P.T.; Collin, D.; Smith, S.B.; Bustamante, C.; Tinoco, I., Jr. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* **2006**, *90*, 250–260. [[CrossRef](#)]
39. McCauley, M.J.; Williams, M.C. Optical tweezers experiments resolve distinct modes of DNA-protein binding. *Biopolymers* **2009**, *91*, 265–282. [[CrossRef](#)]
40. Neuman, K.C.; Nagy, A. Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **2008**, *5*, 491–505. [[CrossRef](#)]
41. Woodside, M.T.; Block, S.M. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Ann. Rev. Biophys.* **2014**, *43*, 19–39. [[CrossRef](#)] [[PubMed](#)]
42. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
43. Markham, N.R.; Zuker, M. UNAFold: Software for nucleic acid folding and hybridization. *Methods Mol. Biol.* **2008**, *453*, 3–31. [[CrossRef](#)]
44. Woodside, M.T.; Anthony, P.C.; Behnke-Parks, W.M.; Larizadeh, K.; Herschlag, D.; Block, S.M. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **2006**, *314*, 1001–1004. [[CrossRef](#)] [[PubMed](#)]
45. Woodside, M.T.; Behnke-Parks, W.M.; Larizadeh, K.; Travers, K.; Herschlag, D.; Block, S.M. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6190–6195. [[CrossRef](#)]
46. Godet, J.; Kenfack, C.; Przybilla, F.; Richert, L.; Duportail, G.; Mely, Y. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res.* **2013**, *41*, 5036–5048. [[CrossRef](#)]
47. Grohman, J.K.; Gorelick, R.J.; Lickwar, C.R.; Lieb, J.D.; Bower, B.D.; Znosko, B.M.; Weeks, K.M. A guanosine-centric mechanism for RNA chaperone function. *Science* **2013**, *340*, 190–195. [[CrossRef](#)]
48. Wang, W.; Naiyer, N.; Mitra, M.; Li, J.; Williams, M.C.; Rouzina, I.; Gorelick, R.J.; Wu, Z.; Musier-Forsyth, K. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res.* **2014**, *42*, 7145–7159. [[CrossRef](#)]
49. Wu, T.; Gorelick, R.J.; Levin, J.G. Selection of fully processed HIV-1 nucleocapsid protein is required for optimal nucleic acid chaperone activity in reverse transcription. *Virus Res.* **2014**, *193*, 52–64. [[CrossRef](#)] [[PubMed](#)]
50. Cen, S.; Khorchid, A.; Gabor, J.; Rong, L.; Wainberg, M.A.; Kleiman, L. Roles of Pr55(gag) and NCp7 in tRNA(3)(Lys) genomic placement and the initiation step of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* **2000**, *74*, 10796–10800. [[CrossRef](#)]
51. Guo, F.; Saadatmand, J.; Niu, M.; Kleiman, L. Roles of Gag and NCp7 in facilitating tRNA(Lys)(3) Annealing to viral RNA in human immunodeficiency virus type 1. *J. Virol.* **2009**, *83*, 8099–8107. [[CrossRef](#)] [[PubMed](#)]
52. Jones, C.P.; Datta, S.A.; Rein, A.; Rouzina, I.; Musier-Forsyth, K. Matrix domain modulates HIV-1 Gag's nucleic acid chaperone activity via inositol phosphate binding. *J. Virol.* **2011**, *85*, 1594–1603. [[CrossRef](#)] [[PubMed](#)]
53. Comas-Garcia, M.; Kroupa, T.; Datta, S.A.; Harvin, D.P.; Hu, W.S.; Rein, A. Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *Elife* **2018**, *7*, e38438. [[CrossRef](#)] [[PubMed](#)]
54. Webb, J.A.; Jones, C.P.; Parent, L.J.; Rouzina, I.; Musier-Forsyth, K. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA* **2013**, *19*, 1078–1088. [[CrossRef](#)] [[PubMed](#)]
55. Williams, M.C.; Rouzina, I.; Wenner, J.R.; Gorelick, R.J.; Musier-Forsyth, K.; Bloomfield, V.A. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6121–6126. [[CrossRef](#)] [[PubMed](#)]
56. Mekdad, H.E.; Boutant, E.; Karnib, H.; Biedma, M.E.; Sharma, K.K.; Malytska, I.; Laumond, G.; Roy, M.; Real, E.; Paillart, J.C.; et al. Characterization of the interaction between the HIV-1 Gag structural polyprotein and the cellular ribosomal protein L7 and its implication in viral nucleic acid remodeling. *Retrovirology* **2016**, *13*, 54. [[CrossRef](#)]

57. Munro, J.B.; Nath, A.; Farber, M.; Datta, S.A.; Rein, A.; Rhoades, E.; Mothes, W. A conformational transition observed in single HIV-1 Gag molecules during in vitro assembly of virus-like particles. *J. Virol.* **2014**, *88*, 3577–3585. [[CrossRef](#)]
58. Dick, R.A.; Zadrozny, K.K.; Xu, C.; Schur, F.K.M.; Lyddon, T.D.; Ricana, C.L.; Wagner, J.M.; Perilla, J.R.; Ganser-Pornillos, B.K.; Johnson, M.C.; et al. Inositol phosphates are assembly co-factors for HIV-1. *Nature* **2018**, *560*, 509–512. [[CrossRef](#)]
59. Todd, G.C.; Duchon, A.; Inlora, J.; Olson, E.D.; Musier-Forsyth, K.; Ono, A. Inhibition of HIV-1 Gag-membrane interactions by specific RNAs. *RNA* **2017**, *23*, 395–405. [[CrossRef](#)]
60. Kenyon, J.C.; Prestwood, L.J.; Lever, A.M. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci. Rep.* **2015**, *5*, 14369. [[CrossRef](#)]
61. Smyth, R.P.; Despons, L.; Huili, G.; Bernacchi, S.; Hijnen, M.; Mak, J.; Jossinet, F.; Weixi, L.; Paillart, J.C.; von Kleist, M.; et al. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods* **2015**, *12*, 866–872. [[CrossRef](#)] [[PubMed](#)]
62. Wilkinson, K.A.; Gorelick, R.J.; Vasa, S.M.; Guex, N.; Rein, A.; Mathews, D.H.; Giddings, M.C.; Weeks, K.M. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **2008**, *6*, e96. [[CrossRef](#)] [[PubMed](#)]
63. Bouchiat, C.; Wang, M.D.; Allemand, J.F.; Strick, T.; Block, S.M.; Croquette, V. Estimating the persistence length of a Worm-Like Chain molecule from Force-Extension measurements. *Biophys. J.* **1999**, *76*, 409–413. [[CrossRef](#)]
64. Odijk, T. Stiff chains and filaments under tension. *Macromolecules* **1995**, *28*, 7016–7018. [[CrossRef](#)]
65. Smith, S.B.; Cui, Y.; Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **1996**, *271*, 795–799. [[CrossRef](#)] [[PubMed](#)]
66. Seol, Y.; Li, J.; Nelson, P.C.; Perkins, T.T.; Betterton, M.D. Elasticity of short DNA molecules: Theory and experiment for contour lengths of 0.6–7 microns. *Biophys. J.* **2007**, *93*, 4360–4373. [[CrossRef](#)] [[PubMed](#)]
67. Wenner, J.R.; Williams, M.C.; Rouzina, I.; Bloomfield, V.A. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.* **2002**, *82*, 3160–3169. [[CrossRef](#)]
68. Baumann, C.G.; Smith, S.B.; Bloomfield, V.A.; Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6185–6190. [[CrossRef](#)]
69. Smith, S.; Finzi, L.; Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **1992**, *258*, 1122–1126. [[CrossRef](#)]
70. McCauley, M.J.; Rouzina, I.; Williams, M.C. Specific Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein Deduced from Hairpin Unfolding. *Methods Mol. Biol.* **2020**, *2106*, 59–88. [[CrossRef](#)]
71. Crooks, G.E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721–2726. [[CrossRef](#)] [[PubMed](#)]
72. Crooks, G.E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **2000**, *61*, 2361–2366. [[CrossRef](#)]
73. Bennett, C.H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268. [[CrossRef](#)]
74. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693. [[CrossRef](#)]
75. Dudko, O.K.; Filippov, A.E.; Klafter, J.; Urbakh, M. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11378–11381. [[CrossRef](#)] [[PubMed](#)]
76. Dudko, O.K.; Hummer, G.; Szabo, A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **2006**, *96*, 108101. [[CrossRef](#)]
77. Dudko, O.K.; Hummer, G.; Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15755–15760. [[CrossRef](#)]

78. Pierse, C.A.; Dudko, O.K. Distinguishing Signatures of Multipathway Conformational Transitions. *Phys. Rev. Lett.* **2017**, *118*, 088101. [[CrossRef](#)]
79. McCauley, M.J.; Rouzina, I.; Williams, M.C. Constructing Free Energy Landscapes of Nucleic Acid Hairpin Unfolding. *Methods Mol. Biol.* **2018**, *1811*, 315–332. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).