Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA

Micah J. McCauley[a], Ioulia Rouzina[b], Kelly A. Manthei[b], Robert J. Gorelick[c], Karin Musier-Forsyth[d], and Mark C. Williams[a,1]

[a]Department of Physics, Northeastern University, Boston, MA 02115; [b]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN 55455; [c]AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21702; and [d]Department of Chemistry and Biochemistry, Center for Retroviral Research and Center for RNA Biology, The Ohio State University, Columbus, OH 43210

Retroviral nucleocapsid (NC) proteins are nucleic acid chaperones that play a key role in the viral life cycle. During reverse transcription, HIV-1 NC facilitates the rearrangement of nucleic acid secondary structure, allowing the transactivation response (TAR) RNA hairpin to be transiently destabilized and annealed to a cDNA hairpin. It is not clear how NC specifically destabilizes TAR RNA but does not strongly destabilize the resulting annealed RNA–DNA hybrid structure, which must be formed for reverse transcription to continue. By combining single-molecule optical tweezers measurements with a quantitative mfold-based model, we characterize the equilibrium TAR stability and unfolding barrier for TAR RNA. Experiments show that adding NC lowers the transition state barrier height while also dramatically shifting the barrier location. Incorporating TAR destabilization by NC into the mfold-based model reveals that a subset of preferential protein binding sites is responsible for the observed changes in the unfolding landscape, including the unusual shift in the transition state. We measure the destabilization induced at these NC binding sites and find that NC preferentially targets TAR RNA by binding to specific sequence contexts that are not present on the final annealed RNA–DNA hybrid structure. Thus, specific binding alters the entire RNA unfolding landscape, resulting in the dramatic destabilization of this specific structure that is required for reverse transcription.

single molecule | force spectroscopy | RNA stretching | RNA binding

The transactivation response (TAR) RNA hairpin is a 59-nt sequence in the long-terminal repeat (LTR) of the HIV-1 genome that forms a 24-bp hairpin (Fig. 1A) (1). This structure is essential in promoting viral transactivator protein (Tat)-mediated transcription. The protein–RNA complex further enhances LTR promoter activity (2). The highly stable TAR hairpin structure that stimulates viral RNA transcription becomes a liability during the early stage of a new infection, as TAR hairpins inhibit the minus-strand transfer step required for reverse transcription (1). To alleviate this inhibition, successful reverse transcription requires a key viral chaperone, the nucleocapsid (NC) protein. In vitro experiments have shown a 3,000-fold stimulation of the rate-limiting step of minus-strand transfer in the presence of NC (3), as NC is required to destabilize TAR RNA and the complementary repeat TAR DNA hairpin to allow subsequent strand annealing (1).

HIV-1 NC is only 55 aa long, consisting of two highly conserved CCHC zinc fingers and a basic N terminus (1) (Fig. 1B). The multiple roles of NC during reverse transcription all use the same "chaperone" activity (1), which describes HIV-1 NC's ability to facilitate the rearrangement of nucleic acids into the most stable structures, with the lowest free energy (1). This chaperone activity is characterized by nucleic acid aggregation, duplex destabilization, and rapid kinetics of protein–nucleic acid interactions (3, 4). Aromatic residues in each zinc finger stack with single-stranded nucleic acid bases, resulting in preferential binding to single-stranded nucleic acids (5). Recent studies suggest that preferred sites for NC-induced destabilization involve guanine-containing base pairs at the boundary with defects in the duplex, which include

mismatches, loops, and bubbles (6–8). However, the contribution of this localized RNA structure destabilization by NC to the facilitation of RNA unfolding and refolding as well as the magnitude of the destabilization remains unclear.

In this work we use single-molecule optical tweezers (OT) to force-unfold the TAR RNA hairpin and characterize the energy landscape of hairpin unfolding. We show that NC dramatically alters the TAR unfolding transition state position and energy. Our results quantify the destabilization induced by specific NC binding to a limited number of paired guanine bases located at the boundaries of TAR stem defects. Such binding effectively creates larger loops in the already interrupted TAR RNA secondary structure, shifting the transition state position and significantly increasing the spontaneous RNA opening probability. In addition to quantifying NC-induced changes to the TAR unfolding landscape, this work provides a previously unidentified case study of a protein that specifically destabilizes particular elements of nucleic acid secondary structure during the unfolding process.

Results

TAR Hairpin Unfolding Is a Two-State Process. To quantify the effect of NC on TAR RNA hairpins, the 59-nt hairpin (Fig. 1A) was ligated to long DNA handles and tethered in a dual-beam optical tweezers apparatus in both the absence and presence of NC (Fig. 1 B and C). For constructs where the TAR hairpin was omitted, cycles of extension and release exhibit the smooth curvature characteristic of DNA handle elasticity (Fig. 1 D and E) over a fixed pulling rate.

Significance

The nucleocapsid protein (NC) of the human immunodeficiency virus type 1 (HIV-1) is critical for HIV-1 replication. NC is required for reverse transcription, in which the viral single-stranded RNA genome is converted into double-stranded DNA for replication in the cell. One critical step of reverse transcription that requires NC involves transient destabilization of the transactivation response (TAR) RNA hairpin into a stable double-stranded RNA–DNA hybrid structure. It is not clear how NC is able to destabilize TAR RNA without destabilizing the resulting double-stranded structure. This work shows for the first time to our knowledge that NC binding to specific defects and sequence contexts in TAR RNA dramatically alters the unfolding landscape, preferentially destabilizing TAR RNA relative to other structures to facilitate reverse transcription.
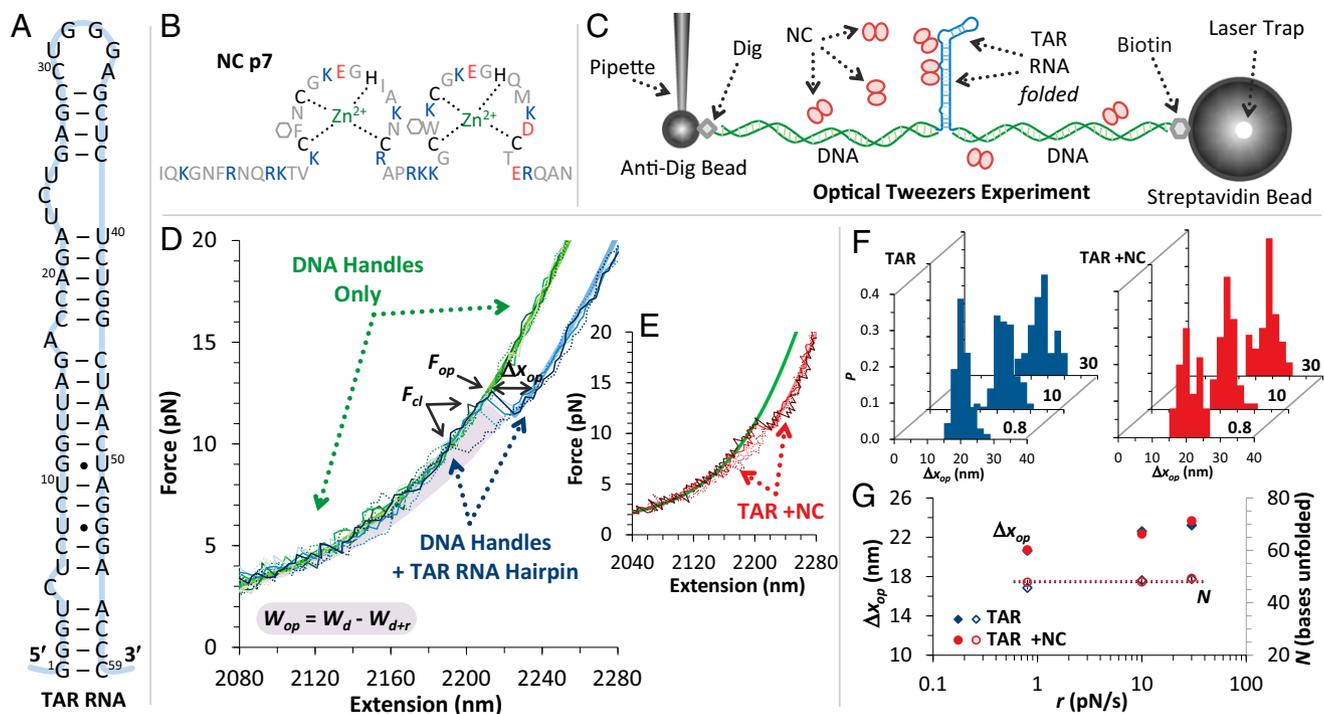

Fig. 1. Probing the interaction of the TAR RNA hairpin and NC. (*A*) The 59-nt sequence and predicted secondary structure of HIV-1 TAR RNA include 24 bp. (*B*) NC consists of two zinc fingers and a basic N terminus. Basic residues are shown in blue and acidic residues in red, and black denotes zinc-coordinating amino acids. Aromatic residues Phe and Trp are marked. (*C*) An optical tweezers experiment tethers a single RNA hairpin between two beads through long DNA handles. The micropipette is translated to increase the tension. (*D* and *E*) Control constructs excluding the hairpin show an elastic response typical of DNA for three cycles of extension/release (solid/dotted green lines). Experimental constructs incorporating TAR RNA hairpins reveal the same elastic response as the DNA handles until interrupted by sudden hairpin opening at ~12 pN (solid/dotted blue lines for TAR and solid/dotted red lines for TAR with NC). Thick solid lines are fits to polymer elasticity models from *SI Materials and Methods*, section 1. Unfolding is characterized by a measured force (F_{op}) and length increase (Δx_{op}), and the hairpin closing force (F_{cl}) is identified where the release and extension data overlap. The shaded region represents the net work done by the instrument to open the hairpin (W_{op}) and represents the energy required to extend the folded construct as the hairpin remains folded (ΔW_d) minus the energy required to extend the handles and the unfolded hairpin construct over the same extension range (ΔW_{d+r}). (*F*) Histograms of measured opening lengths for pulling rates shown in pN/s. (*G*) At each pulling rate the average length (solid symbols) shows a variability, which disappears when corrected for polymer elasticity (open symbols), giving the number of bases unfolded: $n = 47.8 \pm 1.3$ for the TAR hairpin and $N_{NC} = 48.4 \pm 0.5$ in the presence of NC.

When TAR hairpins are incorporated, extension data show discrete length increases due to force-induced hairpin unfolding. The onset of hairpin opening and the completion of hairpin refolding are determined for each cycle (the onset of refolding at very low forces could not be reliably determined).

Unfolding is characterized by a single length increase of ~22 nm for nearly all cycles (Fig. 1*F*). Constant force experiments have shown weak intermediate states for shortened TAR hairpins, but these are not observable in our rapid force-ramp experiments (9). Correcting for the force-dependent elasticity of the stretched construct, as described in *SI Materials and Methods*, section 1, and for the finite width of the folded stem of 2 nm, as seen for other RNA hairpins (10), gives a corrected unfolded hairpin length, *N*. The average number of bases unfolded is 47.8 ± 1.3 bases for TAR alone and 48.4 ± 0.5 bases in the presence of NC (this length is compared with the full hairpin length below). In our experiments, TAR refolding during release, especially at higher force-ramp rates, is much more variable than unfolding during stretching and may not occur as a two-state process (11).

NC Reduces the Equilibrium Free Energy of TAR RNA Hairpin Unfolding. During the force-extension cycle of an OT experiment, the measured work of unfolding (W_{op}) consists of the difference between the integrated work required to extend the folded construct (ΔW_d) minus the work required to extend an unfolded hairpin construct (*SI Materials and Methods*, section 1) over the same extension range (ΔW_{d+r}) (Fig. 1*D*): $W_{op} = \Delta W_d - \Delta W_{d+r}$ (12). An alternate approach considers only the net work performed by

the instrument across the unfolding transition as the sum of the free energy of base pair opening and the entropy required to extend the open RNA hairpin and change the DNA handle extension (13). Both approaches yielded the same result within uncertainty. Similar calculations determine the work obtained through folding, where error introduced by the presence of intermediately folded states should be less than the uncertainty of these experiments. Probability distributions of measured unfolding/folding energies, $P_{op}(W)$ and $P_{cl}(W)$, are shown for varying pulling rates (Fig. 2 *A* and *B*), and the measured work clearly varies with pulling rate. According to the Crooks fluctuation theorem, the intersection of the opening and closing distributions of measured work reflects the free energy of unfolding ($W = \Delta G_o$) (14). This result is verified to be independent of the pulling rate, within uncertainty (Fig. 2*C*). Averaging over all pulling rates gives a final result of $\Delta G_o = 43.3 \pm 0.9\ k_BT$ for TAR unfolding in the absence of NC (matching the equilibrium result above) and $\Delta G_{o,\ NC} = 28.3 \pm 0.9\ k_BT$ in the presence of NC. Alternative methods of deducing the free-energy change (15, 16) are shown in *SI Materials and Methods*, section 2 (Table S1 and Fig. S1), although the key result, the destabilization of the hairpin upon the addition of NC, is the same for all approaches within uncertainty.

NC Shifts the Transition State of Hairpin Unfolding. Force probability distributions $P(F_{op})$ are fitted across all pulling rates to the dynamic force spectroscopy model of Dudko et al. (17), which is described in *SI Materials and Methods*, section 3 (Fig. 3*A* and Figs. S2 and S3). Fits to this model determine rates of hairpin opening in the absence of force (k_{op}^o) and the distance to and height of the
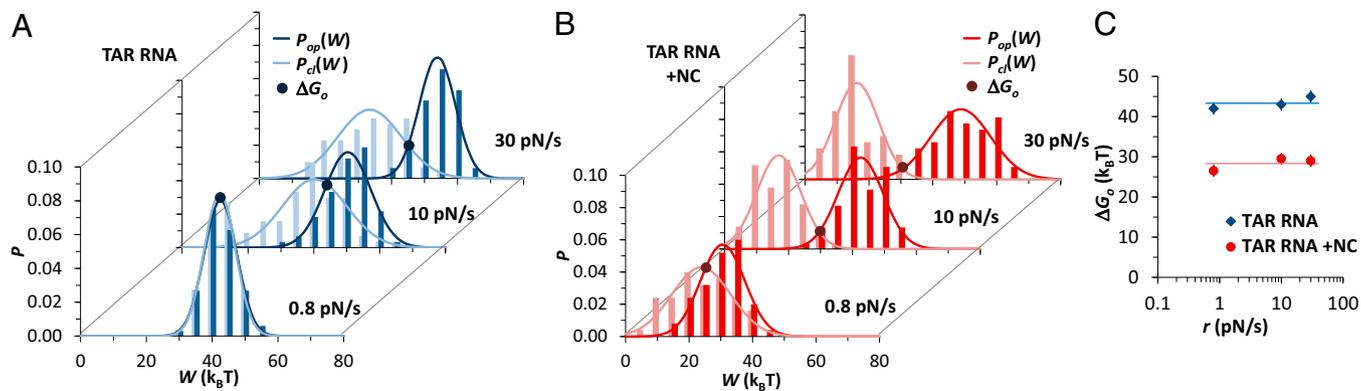
Fig. 2. Equilibrium energies of hairpin opening. (A) Normalized probability densities of measured work during unfolding [$P_{op}(W)$, blue] and closing [$P_{cl}(W)$, cyan] for the TAR RNA hairpin over pulling rates of 30 pN/s, 10 pN/s, and 0.8 pN/s ($n = 250$ opening events). Solid lines are fits to Gaussian distributions to guide the eye. Distributions cross at the equilibrium work ($W = \Delta G_o$), marked by solid circles. (B) Distributions of work done during unfolding (red) and folding (pink) in the presence of NC ($n = 162$ cycles). (C) TAR (blue) and TAR with NC (red) unfolding free energies for distributions shown in A and B. Averaged over all rates, the measured free energy and SE are $\Delta G_o = 43.3 \pm 0.9\ k_BT$ in the absence and $\Delta G_{o,NC} = 28.3 \pm 0.9\ k_BT$ in the presence of NC. Details of the free-energy measurement are discussed in *SI Materials and Methods*, section 2.

transition barrier ($x^\dagger_{op}$ and $\Delta G^\dagger_{op}$). TAR RNA alone exhibits $x^\dagger_{op} = 9.9 \pm 1.1$ nm, corresponding to roughly half of the hairpin opening length, whereas $\Delta G^\dagger_{op} = 27.0 \pm 2.2\ k_BT$ and $k^o_{op} = (8 \pm 5) \times 10^{-9}\ \mathrm{s}^{-1}$. Surprisingly, at the highest pulling rate, addition of NC increases the observed hairpin unfolding force, suggesting initially that NC might stabilize the hairpin (Fig. 3B). However, these effects are associated with a shorter distance from the folded to the transition state, $x^\dagger_{op,NC} = 4.8 \pm 0.6$ nm. In other words, the opening transition rate is less facilitated by force due to the shorter TAR RNA elongation required to reach the transition state in the presence of NC protein. Furthermore, $\Delta G^\dagger_{op,NC} = 14.3 \pm 1.3\ k_BT$ in the presence of NC is reduced to almost half of its value without NC. Finally, the rate of hairpin opening in the absence of force increases ~10,000-fold in the presence of NC to $k^o_{op,NC} = (1.2 \pm 0.8) \times 10^{-4}\ \mathrm{s}^{-1}$. The short transition state distance for TAR is unusual for a long hairpin, and the still shorter distance in the presence of NC is striking.

Complementary analysis of the kinetics based on the cumulative probability of unfolding yielded the unfolding rate as a function of force, $k_{op}(F)$ (18). This is reported in the presence and absence of NC, according to *SI Materials and Methods*, section 4. Force-dependent opening rates for the three pulling rates agree well with each other (Fig. 3C) and can be universally fitted to yield transition state parameters similar to those discussed above within uncertainty (Table S2). The kinetics of hairpin closing, however, do not agree across the pulling rates, likely due to the presence of intermediates (Fig. S4) (19). The main effect of NC is to lower the transition state energy and move it closer to the closed hairpin state, a result well outside the uncertainty of these fits.

mfold Quantifies Destabilization of TAR Due to Specific NC Binding. mfold provides a theoretical estimate of the overall energy of hairpin folding and the energy per base pair for a given
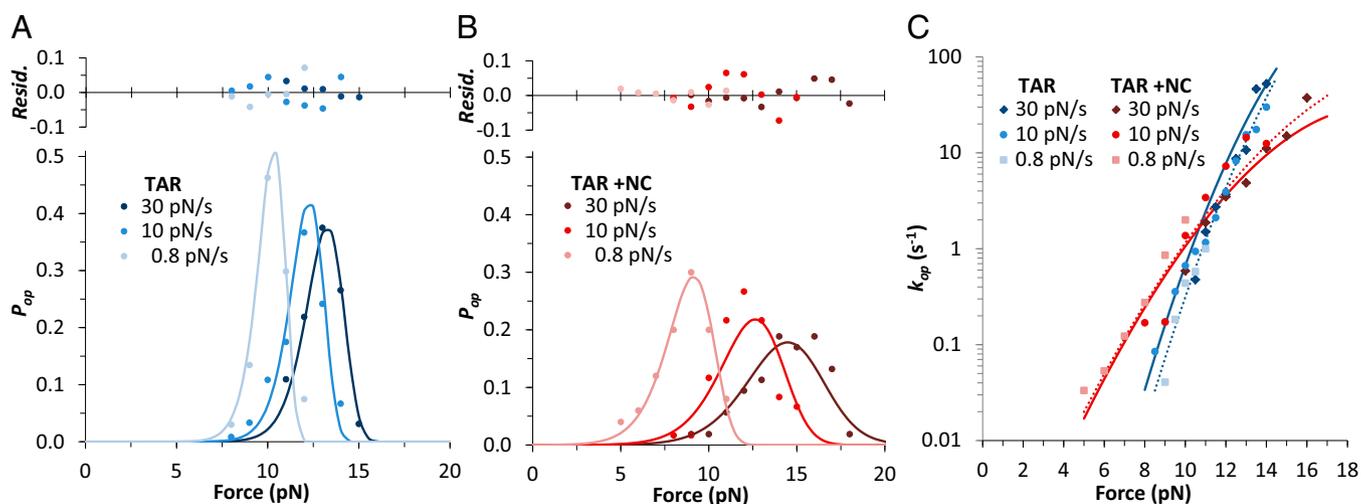
Fig. 3. Kinetics and thermodynamics of TAR unfolding and folding. (A) Normalized TAR RNA opening probabilities $P_{op}(F)$ recast as points for pulling rates of 30 pN/s, 10 pN/s, and 0.8 pN/s (dark blue, blue, and cyan) globally fitted to Eq. S7 (solid lines) described in *SI Materials and Methods*, section 3, and including residuals. Standard histogram bin uncertainties are omitted for clarity, and fits shown are for shape factor $\nu = 0.5$, as discussed in the text. Averaged fitted parameters (with SE) were found: $\Delta x^\dagger_{op} = 9.9 \pm 1.1$ nm, $\Delta G^\dagger_{op} = 27.0 \pm 2.2\ k_BT$, and $k^o_{op} = (8 \pm 5) \times 10^{-9}\ \mathrm{s}^{-1}$ for TAR with $n = 250$ and $\chi^2_\nu \sim 1$ for all fits. (B) In the presence of 50 nM NC, global fits over 30 pN/s, 10 pN/s, and 0.8 pN/s (dark red, red, and pink) yield $\Delta x^\dagger_{op,NC} = 4.8 \pm 0.6$ nm, $\Delta G^\dagger_{op,NC} = 14.3 \pm 1.3\ k_BT$, and $k^o_{op,NC} = (1.2 \pm 0.8) \times 10^{-4}\ \mathrm{s}^{-1}$, where $n = 162$ and $\chi^2_\nu \sim 0.8$ for all fits. Histogram bin widths were scaled for direct comparisons between A and B. (C) TAR RNA hairpin opening rate as a function of unzipping force, $k_{op}(F)$, omitting/including NC, calculated according to Eqs. S9 and S10 in *SI Materials and Methods*, section 4. Solid lines represent the fit to $P_{op}(F)$, and dotted lines are direct fits of $k_{op}(F)$ to Eq. S6. The full list of fitted values is compared in Table S2.
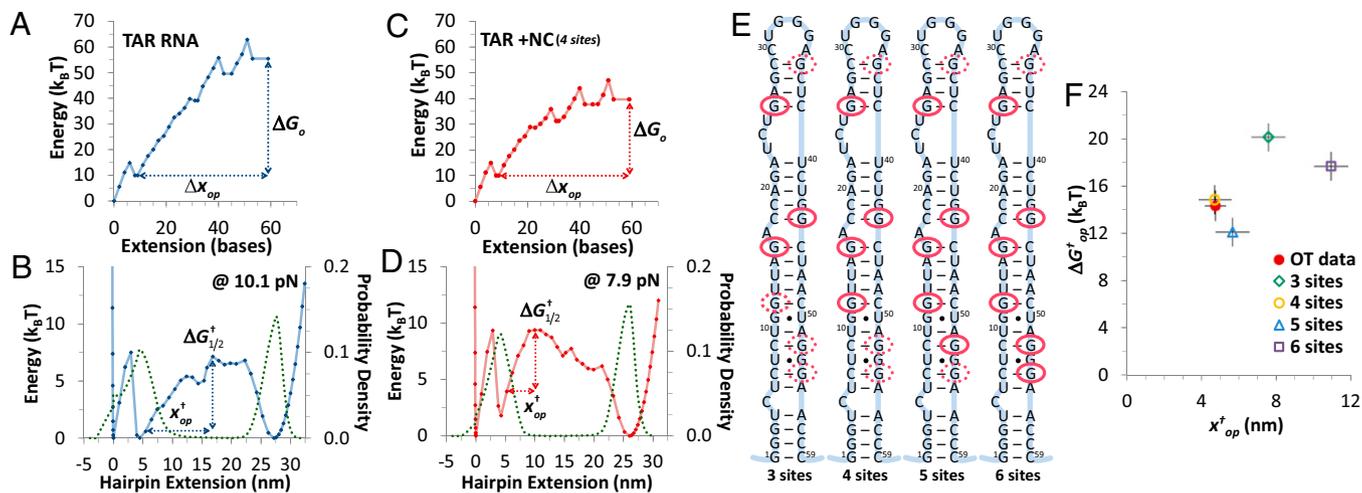
Fig. 4. Transition state predictions for the TAR hairpin in the absence and presence of NC. (*A*) Theoretical energy profiles determined from mfold. Horizontal and vertical lines indicate Δx_{op} and ΔG_o for TAR RNA (blue), for the lowest energy state seen in these experiments, as discussed in the text. (*B*) Free-energy profiles at $F_{1/2}$, where folded and unfolded state free energies are equal, with corrections for ssRNA elasticity and added potentials for open and closed states and probability distributions for the two states (green) (*SI Materials and Methods*, section 5). Horizontal and vertical lines indicate theoretical estimates of $x^{\dagger}_{op}$ and $\Delta G^{\dagger}_{1/2}$ for TAR RNA (blue): $F_{1/2}$ = 10.1 ± 0.1 pN, $x^{\dagger}_{op}$ = 11.3 ± 0.9 nm, and $\Delta G^{\dagger}_{op}$ = 30.7 ± 2.0 k_BT. (*C* and *D*) Predicted destabilization of TAR in the presence of NC (for four bound NCs). The resulting landscapes (red) give $F_{1/2,NC}$ = 7.9 ± 0.1 pN, $x^{\dagger}_{op,NC}$ = 4.7 ± 0.9 nm, and $\Delta G^{\dagger}_{op,NC}$ = 14.8 ± 1.2 k_BT. (*E*) Potential NC binding sites located at defect-adjacent G-containing base pairs are circled in red for four test cases. Sites marked with solid circles were uniformly destabilized by a total δG_o = 16 k_BT. (*F*) Values of the transition state location and height calculated from mfold for each case shown in *E*. A 2D *Z*-test of the data with the various models gives a probability of 0.74 for the four-site model and 0.08 for the five-site model, whereas the three- and six-site models each have probability of less than 0.001.

sequence (20). The free-energy profile for opening n nucleotides of TAR RNA at zero force was calculated as a sum of unzipping free energies for the corresponding elements of the TAR structure, using mfold energies per base pair, $G(n_i, F = 0)$ (Fig. 4*A*). Calculations were performed at standard conditions approximately equivalent to our experimental conditions, as discussed in *SI Materials and Methods*, section 6. In addition to the hairpin unfolding energy, further insight can be derived by subtracting the mechanical work $F_{1/2} \cdot x$ done by the applied force, to find the landscape, $G(x_i, F_{1/2})$, where $F_{1/2}$ is the critical force at which folded and unfolded states have equal free energy (Fig. 4*B*) (10, 21). For the full TAR RNA sequence, $F_{1/2}$ = 10.1 ± 0.1 pN, after correcting for single-stranded RNA (ssRNA) elasticity during unfolding and the 2-nm width of the hairpin stem.

Our experimentally determined value of the unfolding energy for TAR RNA alone, 44 k_BT, is significantly lower than the value calculated in mfold, 56 k_BT. Similarly, only 48 bases of an expected length of 59 bases were observed to open during experimental unfolding studies, as noted above. A consistent explanation for both observations is found in the calculated TAR RNA $G(x_i, F_{1/2})$ profile, which shows a deep free-energy minimum at an extension of ~5 nm corresponding to the lowest C bulge (Fig. 4*B*). We also show the calculated occupancy probability, smoothed to represent the system elasticity. The transition is predicted to be primarily between a partially frayed state and the unfolded state, as seen for other hairpins (10, 21). The energy landscape suggests a separate unfolding step at low extension, which is not observed, although the landscape near the fully folded state may not be precisely represented by the model. Theoretical values of the unfolded and transition states are determined to be ΔG_o = 41.5 ± 1.5 k_BT, $x^{\dagger}_{op}$ = 11.3 ± 0.9 nm, and $\Delta G^{\dagger}_{op}$ = 30.7 ± 2.0 k_BT, where the uncertainties are primarily due to uncertainty in the location of the frayed state. The unfolding length of Δx = 21.8 ± 0.5 nm, determined at $F_{1/2}$, is in good agreement with the unfolding length at the transition force measured in the OT experiments.

Previous studies have shown a preference for NC binding to G residues located adjacent to local defects, such as G·U wobble pairs, bulges, mismatches, and loops (6, 7). All such potential

sites in TAR RNA are circled in Fig. 4*E*. In our model, we reduced the free energy of these potential sites by fixed amounts, $\delta G_{NC}(n_i)$, and observed the resulting free-energy profiles, $G_{NC}(n_i, F = 0)$ and $G_{NC}(x_i, F_{1/2})$. To match the experimental difference measured in OT experiments, the total destabilization was required to match that observed in experiments within uncertainty. Next, we looked for the combination of NC-induced destabilization that would recover the measured properties of the transition state, especially the opening distance to the transition, $x^{\dagger}_{op}$. Spreading evenly the net destabilization over all seven potential sites, $C^7 \cdot G^{54}$, $C^9 \cdot G^{52}$, $G^{12} \cdot C^{49}$, $G^{16} \cdot C^{45}$, $C^{18} \cdot G^{44}$, $G^{26} \cdot C^{39}$, and $C^{29} \cdot G^{36}$, did not lead to the correct position of the transition state. Extensive variation of $\delta G_{NC}(n_i)$ shows the loop-adjacent site $C^{29} \cdot G^{36}$ is not destabilized by NC. Presented in Fig. 4 are selected mfold models with six, five, four, or three sites each destabilized by equal amounts (solid lines denote destabilized sites). Although it is possible that different sites may show varying amounts of destabilization, these details cannot be resolved in our experiments. The energy landscape, $G_{NC}(n_i, F = 0)$ and $G_{NC}(x_i, F_{1/2})$, is determined for each case, and the example of four NC binding sites is shown in Fig. 4 *C* and *D*. For the models in Fig. 4*E*, Fig. 4*F* illustrates the results for the calculated transition state parameters $\Delta G^{\dagger}_{op}$ and $x^{\dagger}_{op}$ from this analysis. The four-binding-site model is the best match to the measured results of the OT experiments within uncertainty, giving the transition state parameters, $x^{\dagger}_{op,NC}$ = 4.7 ± 0.9 nm and $\Delta G^{\dagger}_{op,NC}$ = 14.8 ± 1.2 k_BT, and the equilibrium TAR unfolding free energy in the presence of NC, $\Delta G_{o,NC}$ = 25.6 ± 1.5 k_BT. We find a new transition force, $F_{1/2,NC}$ = 7.9 ± 0.1 pN, and unfolding length Δx_{NC} = 20.8 ± 0.5 nm. Comparisons with experiment are made in Fig. 5 and Table S2 and are discussed below. Although there is a small probability associated with the five-NC binding-sites state model (Fig. 4), the following discussion is not altered by this possibility. Finally, this close match to experiment was found only for this very specific combination of $\delta G_{NC}(n_i)$ values, in which no destabilization occurs in the part of the TAR stem below this new transition state. Thus, NC destabilizes the TAR RNA hairpin by targeting specific sites on that part of the stem, further reducing the distance from the closed state to the transition state.

Discussion

TAR RNA Hairpin Stability Is Weakened by Multiple Duplex Defects.
Comparing the interrupted sequence of the TAR RNA hairpin stem to regular, "unpatterned" DNA hairpins of similar stem and loop length, which were studied extensively by force-unfolding with OT by Woodside et al. (10), we observe that the overall TAR RNA stability is lower, whereas the equilibrium extension change upon TAR unfolding is the same. Several defects along the stem are responsible for this overall decrease in the TAR stability. These defects also lead to a relatively low value of the measured and calculated transition force of $F_{1/2} = 10.1$ pN, which is typical of much shorter hairpin stems only 10 bp in length. A TAR duplex with no defects (~24 bp in the stem) would have an expected $F_{1/2} > 15$ pN. A TAR RNA hairpin, modified by the replacement of the lowest 4 bp and the C bulge with a stabilizing G·C base pair and excluding the bases below the lowest bulge, was shown to be more stable than the wild-type hairpin studied here, with $F_{1/2}$ ~12 pN or higher, depending on solution conditions (11, 22).

In the absence of protein, the TAR hairpin opening transition state is predicted by our landscape calculations to be located at the $G^{21} \cdot C^{41}$ base pair, the second base pair below the UCU bulge. mfold calculations reproduce this TAR opening length to the transition state, which lies about half way to the unfolded state (Fig. 5). However, regular hairpins with unpatterned stems typically have their opening transition state close to the loop, such that for an ~24-bp hairpin with a 6-base loop, the distance from the folded state to the transition state would include nearly

the entire hairpin length less the length of the loop. For TAR, the low stability of the 4-bp helix separating the terminal loop from the nearby UCU bulge means that this entire region effectively behaves as a giant hairpin loop during TAR force-unfolding (10). The long effective size of this loop shifts the opening state to the midpoint between the closed and opened state of the hairpin. This defect must be also responsible for its faster zero-force opening rate of ~10^{-8} s^{-1}, which is typical for 20-bp hairpins (10). In summary, these mfold-based landscape model calculations reproduce and explain all of the essential measured features of the full TAR RNA hairpin.

NC Targets Specific Locations on the TAR RNA Hairpin. Comparing our results for HIV-1 TAR RNA unfolding by force in the absence and presence of NC, we conclude that only a few specifically bound NC molecules alter the TAR RNA opening pathway by destabilizing several sites on this interrupted hairpin. Our data allow us to estimate the destabilization at these sites to be 2.4 kcal/mol (~4.0 k_BT), which implies complete melting of the G–C base pairs shown in Fig. 4, as well as some additional stabilization of the single-stranded state by NC. In general, weakly base-paired G bases were recently shown to be sites for preferential NC binding accompanied by duplex destabilization in 2-aminopurine fluorescence studies (6), in single-molecule FRET studies (23), in SHAPE footprinting studies of HIV-1 NC (8) and MLV NC (24), and in a recent NMR study of MLV NC (25). Our results support this observation while also identifying the specific subset of these sites on TAR that are responsible for altering its unfolding landscape. The observed specific sites of NC-induced duplex destabilization are not the most energetically favorable, which should instead correspond to G-rich single-stranded regions bordered by duplexes (8, 26, 27), such as those found in the major TAR loop. Although there is likely NC binding to the TAR loop, our results show that such binding has negligible effect on TAR stability. The differences in NC–RNA binding and structure destabilization reflect the two independent NC functions in viral RNA: selection and packaging vs. nucleic acid refolding during reverse transcription. The ability of NC to bind selectively to very specific sites during packaging and to also bind to multiple less specific sites to facilitate nucleic acid rearrangement constitutes independent activities of NC that contribute to its functioning at different stages of the virus life cycle.

Specific Binding Drives the Unusual Shift in the TAR Transition State Distance. A reduction of the hairpin unfolding free-energy barrier may be caused by increasing concentrations of NC or by decreasing the concentration of salt in solution. Previous investigations have examined effects on hairpin force-unfolding due to changing solution ionic strength (22, 28), temperature (29), and stabilizing ligand (30). Importantly, in all of these previous studies the location of the opening transition state was not affected by the variation of solution conditions or the addition of ligand. However, the overall unfolding free energy, the transition state free energy, and the transition kinetics were all affected in these experiments. Solution factors that uniformly affect the stability of all nucleic acid structural elements generally leave the position of the transition state unchanged, although position shifts due to multistate unfolding have been observed in multistem hairpins (31). In contrast, NC binding to TAR alters the two-state transition pathway, specifically shortening the transition state distance, and this indicates that NC's effect on the stability of structural elements of TAR must be selective. NC binds to guanines adjacent to unstable stem regions, such as mismatches, loops, and bulges. Binding further destabilizes these weak base pairs, leading to a longer destabilized region running from the apical loop down to the wobble base pair beyond the new transition site (G^{11}–C^{50}). By itself this destabilization is insufficient to induce full duplex unfolding, but can lead to opening of the base pairs destabilized by nearby duplex imperfections. Therefore, in the presence of NC unfolding requires only the
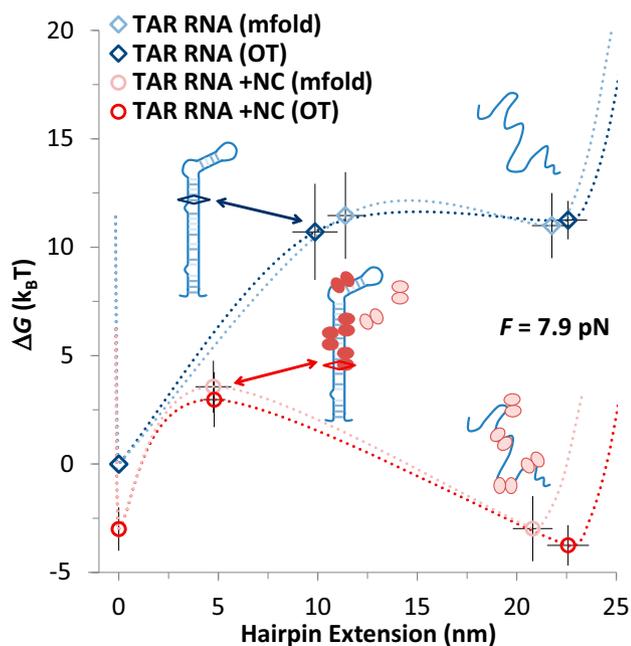
Fig. 5. Specific binding of NC transforms the energy landscape of the TAR RNA hairpin. Shown is a summary of experimental and computational results for TAR RNA hairpin force-unfolding, combining the overall opening length and unfolding free energy with the transition state location into a unified free-energy landscape. Theory and experiment are compared in the absence (cyan/blue) and the presence (pink/red) of 50 nM NC and presented at a common external force of $F_{1/2,NC} = 7.7$ pN. Lines are interpolations to guide the eye. Transition state distances and energies are shown relative to a folded state where the bottom part of the stem is frayed as described in the text. The zero-force TAR and TAR + NC data are offset by a qualitative preference of NC for ssRNA (*SI Materials and Methods*, section 9). Diamonds locate the transition state for TAR RNA and for TAR in the presence of NC (ovals highlight binding sites). NC dramatically enhances the force-free rate of TAR unzipping by specifically destabilizing the interrupted upper part of the TAR stem, thereby shortening the region at the bottom of the TAR hairpin that has to open before reaching the transition state.

destabilization of the lowest part of the stem structure up to the location of the transition state, whereas the net elongation during the transition is still the full TAR RNA length. This model quantitatively elucidates the biophysical mechanism responsible for the experimental observation of a shortened transition state but an unchanged overall elongation upon unfolding.

Conclusions

These experiments form an important novel case study of a protein that locally destabilizes nucleic acid secondary structures at specific locations. HIV-1 NC preferentially destabilizes specific G-containing base pairs adjacent to defects in the secondary structure, thereby leading to further RNA structure fragmentation. We have shown that the effects are particularly significant for TAR as the hairpin already contains several destabilized regions, which are enhanced by the binding of NC. Thus, NC facilitates TAR RNA annealing to its cDNA hairpin by further fragmentation of the interrupted TAR hairpin stem, dramatically shifting the transition state and leading to a shorter critical unfolding helix, while having little effect on the stability of the final annealed long nucleic acid duplex. Our measured change in the unfolding free-energy barrier height as well as shift in position of the transition state drives a 10^4 increase in the zero-force rate of unfolding, which we determine here. Unlike ligands or solution conditions that uniformly alter duplex stability, NC's targeted, structure- and sequence-specific activity appears optimal for the purpose of facilitating nucleic acid rearrangements into the lowest free-energy conformation. This activity is known to be critical for reverse transcription in retroviruses. By identifying the locations and magnitude of TAR RNA destabilization by NC, as well as quantifying the resulting unfolding landscape, the present study adds important insight into the conceptual picture of the molecular mechanism of NC's nucleic acid chaperone activity.

Materials and Methods

Both control and TAR hairpin constructs were tethered between labeled beads as shown in Fig. 1C. See *SI Materials and Methods*, section 7 and Fig. S5 for construct preparation. A 2.1-μm diameter anti–digoxigenin-coated bead (Spherotech) was fixed onto a micropipette tip (WPI), while a 5.4-μm diameter streptavidin-coated bead (Bangs Labs) was held in a dual laser (Lumics) optical trap, described previously (32). Constructs were extended using a subnanometer resolution piezoelectric transducer (nPoint), and forces were recorded on lateral effect detectors (SpotOn; National Instruments). Experiments were performed at constant pulling rates, which are equivalent to force-ramp rate experiments, as discussed in *SI Materials and Methods*, section 8. Chosen pulling rates (0.8 pN/s, 10 pN/s, and 30 pN/s) span the tradeoff between detector noise, piezoelectric stage feedback, and instrument baseline stability. Experimental buffer was 10 mM Hepes, pH 7.5, and 100 mM Na$^+$, and experimental temperature was 23 °C. Both the elasticity of the DNA handles and unfolded RNA segments are characterized by models of polymer elasticity (*SI Materials and Methods*, section 1). Hairpins were also extended after exposure to a solution of NC prepared as described previously (33). The specific binding sites on TAR RNA appear to be saturated at 50 nM NC (34, 35), as discussed further in *SI Materials and Methods*, section 9.

1. Levin JG, Guo J, Rouzina I, Musier-Forsyth K (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: Critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* 80:217–286.
2. Berkhout B, Jeang KT (1992) Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. *J Virol* 66(1):139–149.
3. Vo MN, Barany G, Rouzina I, Musier-Forsyth K (2006) Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol* 363(1):244–261.
4. Cruceanu M, Gorelick RJ, Musier-Forsyth K, Rouzina I, Williams MC (2006) Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol* 363(5):867–877.
5. Wu H, et al. (2013) Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res* 171(2):263–277.
6. Godet J, et al. (2013) 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* 41(9):5036–5048.
7. Grohman JK, et al. (2013) A guanosine-centric mechanism for RNA chaperone function. *Science* 340(6129):190–195.
8. Wilkinson KA, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6(4):e96.
9. Hyeon C, Thirumalai D (2007) Mechanical unfolding of RNA: From hairpins to structures with internal multiloops. *Biophys J* 92(3):731–743.
10. Woodside MT, et al. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci USA* 103(16):6190–6195.
11. Li PT, Collin D, Smith SB, Bustamante C, Tinoco I, Jr (2006) Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys J* 90(1):250–260.
12. Greenleaf WJ, Frieda KL, Foster DAN, Woodside MT, Block SM (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science* 319(5863):630–633.
13. Collin D, et al. (2005) Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* 437(7056):231–234.
14. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 60(3):2721–2726.
15. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78:2690–2693.
16. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. *J Comput Phys* 22:245–268.
17. Dudko OK, Hummer G, Szabo A (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys Rev Lett* 96(10):108101.
18. Dudko OK, Hummer G, Szabo A (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci USA* 105(41):15755–15760.
19. Zhang Y, Dudko OK (2013) A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc Natl Acad Sci USA* 110(41):16432–16437.
20. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415.
21. Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001–1004.
22. Vieregg J, Cheng W, Bustamante C, Tinoco I, Jr (2007) Measurement of the effect of monovalent cations on RNA hairpin stability. *J Am Chem Soc* 129(48):14966–14973.
23. Cosa G, et al. (2004) Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys J* 87(4):2759–2767.
24. Gherghe C, et al. (2010) Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci USA* 107(45):19248–19253.
25. Miller SB, Yildiz FZ, Lo JA, Wang B, D'Souza VM (2014) A structure-based mechanism for tRNA and retroviral RNA remodelling during primer annealing. *Nature* 515(7528):591–595.
26. Bazzi A, et al. (2012) Intrinsic nucleic acid dynamics modulates HIV-1 nucleocapsid protein binding to its targets. *PLoS One* 7(6):e38905.
27. Vuilleumier C, et al. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: A fluorescence study. *Biochemistry* 38(51):16816–16825.
28. Bizarro CV, Alemany A, Ritort F (2012) Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods. *Nucleic Acids Res* 40(14):6922–6935.
29. Stephenson W, et al. (2014) Combining temperature and force to study folding of an RNA hairpin. *Phys Chem Chem Phys* 16(3):906–917.
30. Anthony PC, Perez CF, García-García C, Block SM (2012) Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proc Natl Acad Sci USA* 109(5):1485–1489.
31. Liphardt J, Onoa B, Smith SB, Tinoco I, Jr, Bustamante C (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science* 292(5517):733–737.
32. Chaurasiya KR, Paramanathan T, McCauley MJ, Williams MC (2010) Biophysical characterization of DNA binding from single molecule force measurements. *Phys Life Rev* 7(3):299–341.
33. Wu W, et al. (1996) Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. *J Virol* 70(7):7132–7142.
34. Webb JA, Jones CP, Parent LJ, Rouzina I, Musier-Forsyth K (2013) Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA* 19(8):1078–1088.
35. Jones CP, Datta SA, Rein A, Rouzina I, Musier-Forsyth K (2011) Matrix domain modulates HIV-1 Gag's nucleic acid chaperone activity via inositol phosphate binding. *J Virol* 85(4):1594–1603.