RNA adapts its flexibility to efficiently fold and resist unfolding

Sukjin S. Jang[1,†,§], Korak Kumar Ray[1,‡,§], David G. Lynall[2,§], Kenneth L. Shepard[2], Colin Nuckolls[1,*], Ruben L. Gonzalez, Jr.[1,*]

[1] *Department of Chemistry, Columbia University, New York, NY 10027 USA*

[2] *Department of Electrical Engineering, Columbia University, New York, NY 10027 USA*

[†] *Current Address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115 USA*

[‡] *Current Address: MRC-Laboratory of Medical Sciences, London W12 0HS, UK*

[§] These authors contributed equally to this work.

* To whom correspondence may be addressed: Colin Nuckolls, Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, Tel: (212) 854-6289, Email: cn37@columbia.edu.; Ruben L. Gonzalez, Jr., Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, Tel: (212) 854-1096, Email: rlg2118@columbia.edu

Abstract

Recent studies have demonstrated that the mechanisms through which biopolymers like RNA interconvert between multiple folded structures are critical for their cellular functions. A major obstacle to elucidating these mechanisms is the lack of experimental approaches that can resolve these interconversions between functionally relevant biomolecular structures. Here, we dissect the complete set of structural rearrangements executed by an ultra-stable RNA, the UUCG stem-loop, at the single-molecule level using a nano-electronic device with microsecond time resolution. We show that the stem-loop samples at least four conformations along two folding pathways leading to two distinct folded structures, only one of which has been previously observed. By modulating its flexibility, the stem-loop can adaptively select between these pathways, enabling it to both fold rapidly and resist unfolding. This paradigm of stabilization through compensatory changes in flexibility broadens our understanding of stable RNA structures and is expected to serve as a general strategy employed by all biopolymers.

Main Text

Introduction

The unique ability of biopolymers like nucleic acids and proteins to fold into and stably maintain complex three-dimensional structures is essential for their biological functions [1-5]. However, biopolymers rarely exist as a single static structure over the course of their functional lifetimes [6-8]. Instead, they form an ensemble of interconverting conformations with varying stabilities [3, 4, 9]. To understand the functions of biomolecules, it is essential to elucidate how these interconverting conformations contribute to the stability and dynamics of the ensemble as a whole.

In particular, the functional dynamics of nucleic acids, most prominently RNA, are governed by chemical properties distinct from proteins and other biomolecules. The four ribonucleotide building blocks can form stable short-range interactions in a variety of combinations

that collectively yield a wide range of possible RNA secondary structures [10]. Formation of a specific secondary structure further sets the stage for establishing the longer-range tertiary interactions typically responsible for the three-dimensional organization of the structured RNA [1, 3, 11]. Thus, elucidating the conformational dynamics of these foundational RNA secondary structural elements is essential for understanding the functional dynamics of larger-scale RNA structures.

One such foundational RNA secondary structural element is a stem-loop, consisting of a double-stranded stem capped by a single-stranded loop. At least two families of stem-loops, in which the loop of each family is composed of a specific subset of four-nucleotide 'tetraloop' sequences (Figure 1), have been shown to be exceptionally stable and widely occurring [12-14]. Due to their prevalence and unanticipated stabilities, such ultra-stable stem-loops are thought to serve as nucleation sites for RNA folding, as elementary building blocks that anchor larger and more complex RNA structures, and as binding sites for RNA-binding proteins [15-18]. Despite decades of work investigating these ultra-stable RNA tetraloops [12, 13, 19-28], however, the unique properties of these specific tetraloop sequences that allow them to confer exceptional stability to stem-loops over other loop sequences remain unknown.

Structural studies suggest this stability arises from the ability of these tetraloops to form unique loop structures that establish a network of non-canonical interactions. For example, the ultra-stable UUCG stem-loop (containing the UUCG tetraloop sequence) has been shown to adopt a loop structure that forms a trans-G-U wobble pair between G_{L4} and U_{L1} and a stacking interaction between U_{L1} and C_{L3} (Figure 1a) [13, 28-30]. Ensemble fast kinetics studies, however, suggest that at least some of this stability arises from relatively long-lived intermediates that facilitate folding [19-21, 31]. Unfortunately, it is challenging to determine the structural identities of these intermediates from the ensemble-averaged data obtained in these studies, making it difficult to elucidate the mechanism through which these intermediates contribute to the stability. Moreover, computational studies predict that stem-loops do not exclusively fold into the single

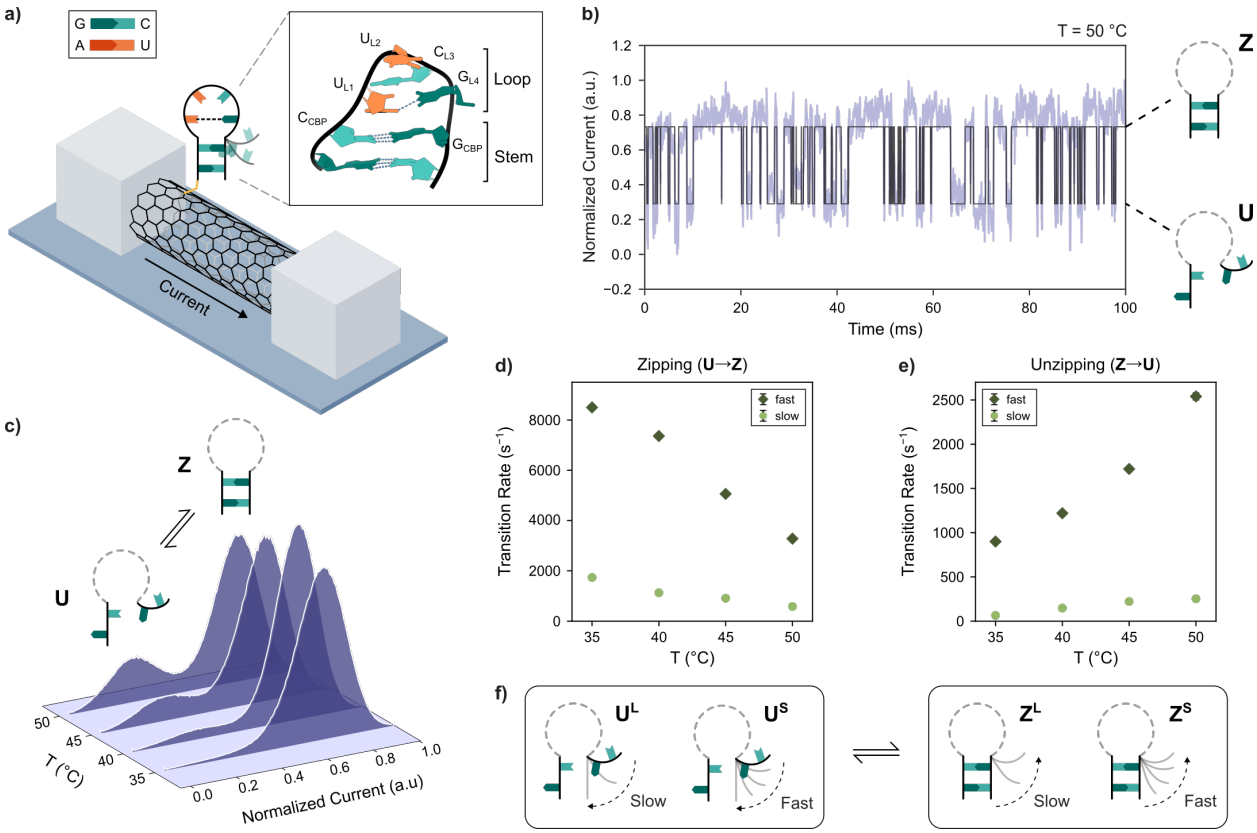

Figure 1. Following the zipping and unzipping kinetics of a single RNA stem-loop using smFET. (a) *A schematic of the smFET experimental platform consisting of a single-walled carbon nanotube connected between two electrodes, with the UUCG stem-loop tethered to the carbon nanotube surface. The inset details the three-dimensional structure of the stem-loop construct (PDB ID: 1F7Y) highlighting the stem, the loop, and the nucleotides forming the loop and the closing base-pair.* ***(b)*** *A 100 ms representative observation window of a normalized current vs. time trajectory showing the real-time zipping dynamics of a single UUCG stem-loop as it transitions between* ***U*** *and* ***Z*** *conformations at 50°C. The raw signal is shown in purple and the idealized trajectory in black.* ***(c)*** *Histograms of normalized current across an entire 60 s trajectory collected at 35 °C, 40 °C, 45 °C, and 50 °C.* ***(d, e)*** *The transition rates for* ***(d)*** *zipping and* ***(e)*** *unzipping of the stem-loop at varying temperatures.* ***(f)*** *A schematic of the observed slow and fast kinetic phases, and the corresponding conformations that they transition from.*

static structures captured in structural studies, but instead exist as a dynamic ensemble of closely related interconverting folded conformations [24, 25, 27, 32]. The net stability of such an ensemble would then arise collectively, from the number of populated conformations, their individual thermodynamic stabilities, and their rates of interconversion. Given these various perspectives, a single unifying model defining the exceptional stability of stem-loops has proven challenging to elucidate.

In this work, we have characterized the complete conformational ensemble of the UUCG stem-loop. Using a carbon nanotube-based nano-electronic device combined with a precise

temperature-control setup, we have recorded the dynamics of a single UUCG stem-loop at microsecond resolution over a range of experimental temperatures (Figure 1). We have identified the two distinct conformations the stem-loop folds into, and resolved the differences in the loop structures adopted by these conformations, that result in varying flexibilities of the folded stem-loop. Only one of these folded conformations, the more rigid one, has been previously identified as the 'native' folded conformation for this motif [13, 28, 30]. We demonstrate that the non-canonical interactions in the native loop structure stabilize this native folded conformation by increasing the rate of structure formation. On the other hand, we show that when faced with destabilizing perturbations, it is the ability of the stem-loop to sample the more flexible 'alternative' loop structure discovered here while maintaining a zipped stem that allows it to retain its overall folded structure and resist unfolding. Beyond underpinning the physical basis for the exceptional stability of the UUCG stem-loop, this paradigm of stabilization through compensatory changes in flexibility represents a fundamental shift in our understanding of RNA folding, structure, dynamics, and function. We expect this paradigm to be a general strategy employed in the folding of other structured biopolymers. Thus, this conceptual breakthrough will be crucial in the design of mutations and small-molecule ligands to both disrupt [33] and programmatically engineer [16] specific biomolecular functions.

Results

The UUCG stem-loop exists as an ensemble of two zipped and two unzipped conformations

We utilized a single-molecule field effect transistor (smFET) device to investigate the dynamics of a model stem-loop construct comprising a stem of two GC base pairs and a UUCG tetraloop [20, 23, 25, 31] (Figure 1a and Table S1; see Material and Methods) over a range of temperatures. Upon tethering the stem-loop to the smFET device [34, 35] (Figures 1a and S1; see Materials and Methods), we observed a current signal that transitioned between a high-current state (normalized current of 0.7 arbitrary units (a.u.)) and a low-current state (normalized current of 0.3 a.u.) (Figure

S2). Based on previous smFET studies [36-39], we could assign the 0.7 and 0.3 a.u. states to zipped and unzipped stem-loop conformations, $\boldsymbol{Z}$ and $\boldsymbol{U}$, respectively (Figure 1b). Consistent with these assignments, the population of $\boldsymbol{U}$ increased as a function of temperature, reflecting the greater propensity of the stem-loop to be unzipped at higher temperatures (Figure 1c). To confirm these assignments, we analyzed the equilibrium between these conformations using a two-state assumption (*i.e.*, making the simplifying assumption that each state corresponded to a single conformation). This analysis accurately recapitulated the thermodynamic stability of this stem-loop as obtained from the standard two-state analysis of a conventional ensemble thermal melting experiment of the same construct (Figures S2-4 and Tables S2 and S3; see Supporting Information for a detailed description of this analysis). In addition, the populations of the two current states at each temperature were in striking agreement with the expected populations for a single stem-loop, demonstrating that the smFET device was indeed reporting on the behavior of a single, tethered stem-loop construct (Figure S4 and Table S3; see Supporting Information for a detailed discussion of this comparison). Collectively, these results validated our assignment of the two current states and demonstrated that the observed zipping dynamics originated from a single stem-loop construct that had been tethered to the smFET device.

Although the two-state thermodynamic analysis described above confirmed that our smFET experiments elicited the same stem-loop stability that is obtained from ensemble experiments, the kinetics of both the zipping and unzipping transitions in the smFET experiments could not be explained by interconversions between single zipped and unzipped conformations. Specifically, the kinetics of both the zipping and unzipping transitions consisted of at least two distinct kinetic phases at all temperatures tested (Figures 1d-e, S5 and S6, and Tables S4 and S5). The slow and fast phases of zipping transitions from $\boldsymbol{U}$ to $\boldsymbol{Z}$ arise from long- and short-lived unzipped conformations, $\boldsymbol{U^L}$ and $\boldsymbol{U^S}$, respectively (Figures 1f and S7a). Similarly, the slow and fast unzipping transitions from $\boldsymbol{Z}$ to $\boldsymbol{U}$ arise from long- and short-lived zipped conformations, $\boldsymbol{Z^L}$

and $\boldsymbol{Z^S}$, respectively (Figures 1f and S7b). Whereas the zipping transition rates decreased with increasing temperature, the unzipping transition rates increased (Figure 1d-e). This is a common characteristic of the folding kinetics of structured biomolecules [40-44].

The fact that these phases could be detected for both $\boldsymbol{U}$ and $\boldsymbol{Z}$ suggests that the rates of interconversion between $\boldsymbol{U^L}$ and $\boldsymbol{U^S}$ and between $\boldsymbol{Z^L}$ and $\boldsymbol{Z^S}$ must be slow compared to our 40 μs experimental acquisition time. The UUCG stem-loop thus exists as an ensemble of at least four conformations, two unzipped, $\boldsymbol{U^L}$ and $\boldsymbol{U^S}$, and two zipped, $\boldsymbol{Z^L}$ and $\boldsymbol{Z^S}$ (Figure 1f). However, because the normalized current values of $\boldsymbol{U^L}$ and $\boldsymbol{U^S}$, as well as $\boldsymbol{Z^L}$ and $\boldsymbol{Z^S}$, could not be distinguished, we were unable to identify transitions between the unzipped conformations or between the zipped conformations using our current analysis tools. This means we could not directly determine how these states were connected. For example, we could not ascertain whether $\boldsymbol{U^L}$ transitions specifically to one of the $\boldsymbol{Z}$ conformations or to both, and the same applies for $\boldsymbol{U^S}$ transitions.

The UUCG stem-loop folds through two distinct zipping pathways

To determine the connectivity between the four observed conformations, we considered two competing kinetic schemes, one where $\boldsymbol{U^L}$ predominately exchanges with $\boldsymbol{Z^L}$, and $\boldsymbol{U^S}$ with $\boldsymbol{Z^S}$ (Figures 2a and S7c), and one where $\boldsymbol{U^S}$ mainly exchanges with $\boldsymbol{Z^L}$, and $\boldsymbol{U^L}$ with $\boldsymbol{Z^S}$ (Figures 2b and S7d). These schemes together encompass all of the possible zipping and unzipping pathways the stem-loop may take (hereafter referred to simply as zipping pathways). We then took advantage of the fact that in each kinetic scheme, the net equilibrium change in thermodynamic parameters between two arbitrary states should be the same regardless of the path used to calculate them. For instance, the net equilibrium change in enthalpy (ΔH_{eq}) between two conformations (*e.g.*, $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$), which represents changes in stabilizing interactions, should

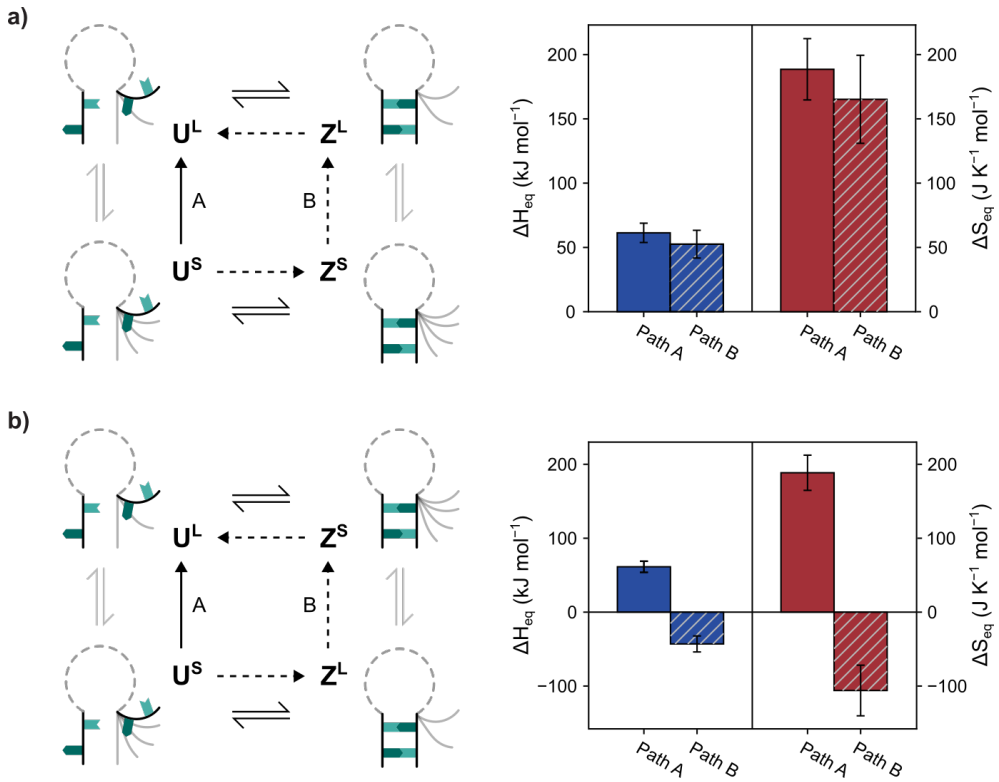

Figure 2. Two possible kinetic schemes for zipping and unzipping of the RNA stem-loop. *The hypothetical kinetic schemes (left) showing predominant transitions* ***(a)*** *between* $\boldsymbol{U^L}$ *and* $\boldsymbol{Z^L}$*, and* $\boldsymbol{U^S}$ *and* $\boldsymbol{Z^S}$*, and* ***(b)*** *between* $\boldsymbol{U^L}$ *and* $\boldsymbol{Z^S}$*, and* $\boldsymbol{U^S}$ *and* $\boldsymbol{Z^L}$*, and the corresponding thermodynamic cycle depicting the direct Path A (solid arrow) and the indirect Path B (dashed arrows) between* $\boldsymbol{U^S}$ *and* $\boldsymbol{U^L}$*. The undetected transitions between the* $\boldsymbol{U}$ *conformations and* $\boldsymbol{Z}$ *conformations are shown with grey equilibrium arrows. The bar graphs (right) depict the calculated* $\boldsymbol{\Delta H_{eq}}$*s (blue) and* $\boldsymbol{\Delta S_{eq}}$*s (red) for these paths for both* ***(a)*** *and* ***(b)****. The error bars represent the propagated errors of fit for the linear fits of free energies as a function of temperature (Figure S8).*

be equal when calculated along a direct path (Path A, solid arrow in Figure 2) or an indirect path (Path B, dashed arrows in Figure 2). Similarly the corresponding equilibrium change in entropy (ΔS_{eq}), which represents changes in the flexibility of the biomolecule and/or the ordering of the solvent, should be equal when calculated along Paths A and B. We calculated the individual ΔH_{eq} and ΔS_{eq} parameters between all possible pairs of conformations using our temperature-dependent measurements of zipping and unzipping kinetics (Figure S8, and Tables S6 and S7; see Materials and Methods for detailed descriptions of these calculations). We then used these parameters to calculate the net equilibrium change in thermodynamic parameters along the direct and indirect paths for the two kinetic schemes. We saw that the net changes for Paths A and B

were equal only for the first scheme and not the second (Figures 2 and S7c-d, a detailed description of this comparison is provided in the Supporting Information). The zipping dynamics observed in our experiments thus overwhelmingly follow the first kinetic scheme.

UUCG stem-loop zipping is therefore characterized by two distinct kinetic pathways that exist in parallel to one another, a slow pathway consisting of transitions between $\boldsymbol{U^L}$ and $\boldsymbol{Z^L}$, and a fast pathway consisting of transitions between $\boldsymbol{U^S}$ and $\boldsymbol{Z^S}$ (Figure 2a). In both pathways, zipping transitions occur at a faster rate than unzipping transitions. Thus, within the temperature range of our experiments, which is below the stem-loop melting temperature, T_m, of 57.0 °C, the equilibria for both pathways favor the zipped conformations over the corresponding unzipped conformations. Notably, upon an increase in temperature, the slow pathway becomes progressively more populated over the fast pathway (Tables S4 and S5), revealing that this zipping pathway of transitions between $\boldsymbol{U^L}$ and $\boldsymbol{Z^L}$ is more dominant at higher temperatures.

The two zipping pathways arise from differences in the structure of the loop

We next sought to identify what RNA structures correspond to $\boldsymbol{U^L}$, $\boldsymbol{U^S}$, $\boldsymbol{Z^L}$, and $\boldsymbol{Z^S}$. We expected our experimentally determined ΔH_{eq} s (Figure S8) to predominantly report on the relative stabilizing interactions present in these conformations, and the ΔS_{eq} s (Figure S8) to predominantly report on their conformational flexibilities. $\boldsymbol{U^L}$, in which the stem-loop possesses the least stabilizing interactions and is the most flexible, thus corresponds to the fully unfolded structure (Figure 3a). Analogously, $\boldsymbol{Z^S}$, in which the stem-loop forms the most stabilizing interactions while also being the most rigid, corresponds to the previously characterized fully folded, native stem-loop structure (Figure 3a).

To determine the identity of $\boldsymbol{Z^L}$, we compared the two zipped conformations. The stem-loop has less stabilizing interactions and is more flexible in $\boldsymbol{Z^L}$ than in $\boldsymbol{Z^S}$ (Figure 3a). Since $\boldsymbol{Z^L}$ and $\boldsymbol{Z^S}$ both correspond to zipped conformations, the RNA must have similar stem structures in

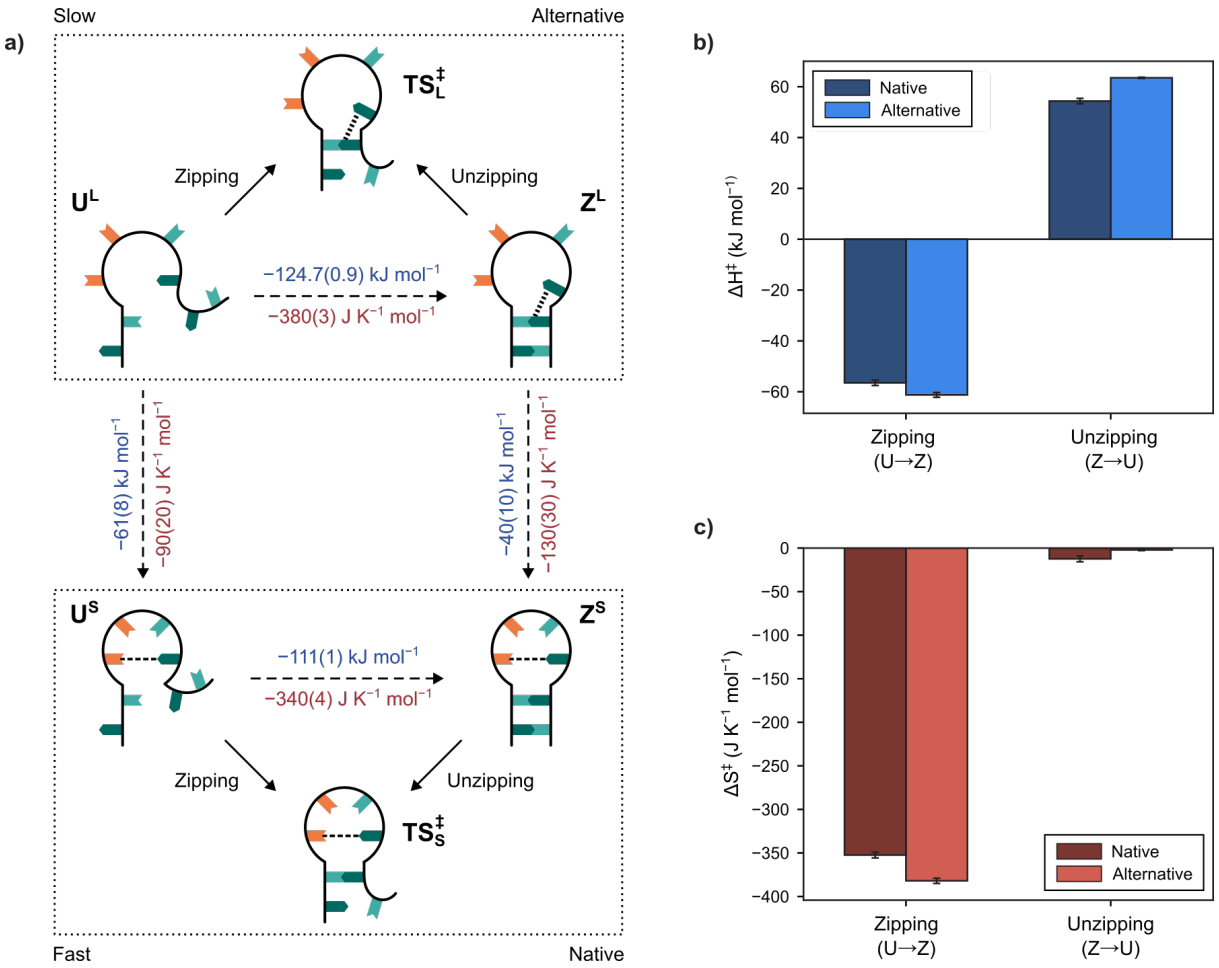

Figure 3. The structural identities of the U and Z conformations. (A) *Schematic represenations of the unzipped conformations (U), transition states ($TS^{\ddagger}$), and zipped conformations (Z) along the alternative (above) and native (below) zipping pathways. The calculated ΔH_{eq}s (blue) and ΔS_{eq}s (red) between each of the observed U and Z conformations are shown.* ***(B)*** *Activation enthalpies ($\Delta H^{\ddagger}$s) for the native (dark blue) and alternative (light blue) pathways.* ***(C)*** *Activation entropies ($\Delta S^{\ddagger}$s) for the native (dark red) and alternative (light red) pathways. The errors (in parentheses) and error bars represent the propagated errors of fit for the linear fits of free energies as a function of temperature (Figure S8).*

both; it is therefore the structure of the UUCG loop that differs between the two conformations. Having established that this loop adopts the previously characterized native structure in Z^S [28, 30], the loop in Z^L must adopt an alternatively folded structure that lacks some of the native Z^S stabilizing interactions (*e.g.*, the *trans*-G_{L4}-U_{L1} wobble pair and/or the stacking interaction between U_{L1} and C_{L3}) and is therefore more flexible. The absence of both of these interactions easily accounts for the ~40 kJ mol^{-1} difference in stabilizing interactions that we observe between the two Z conformations. While the UUCG loop has previously been predicted to be capable of

adopting multiple non-native structures [25, 45, 46], this alternatively folded loop structure has not previously been experimentally observed.

To determine the identity of $\boldsymbol{U^S}$, we similarly compared the two unzipped conformations. $\boldsymbol{U^S}$ possesses more stabilizing interactions and is significantly more constrained than $\boldsymbol{U^L}$ (Figure 3a). Since $\boldsymbol{U^S}$and $\boldsymbol{U^L}$ both represent unzipped conformations, these differences must also arise from differences in the structures of the loop in $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$. Since $\boldsymbol{U^S}$ directly transitions to $\boldsymbol{Z^S}$, we hypothesize that $\boldsymbol{U^S}$ corresponds to a partially unfolded, frayed structure in which the stem-loop is unzipped, but the loop retains its natively folded structure (Figure 3a). Persistence of the native loop structure in $\boldsymbol{U^S}$ is in good agreement with the ~61 kJ mol^{-1} difference in stabilizing interactions we observe between $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$. Our results therefore reveal a hierarchical mechanism of UUCG stem-loop folding in which the loop structure dictates whether the stem-loop folds along a 'native' zipping pathway from $\boldsymbol{U^S}$ to the natively folded $\boldsymbol{Z^S}$ or an 'alternative' zipping pathway from $\boldsymbol{U^L}$ to the alternatively folded $\boldsymbol{Z^L}$.

The alternative UUCG loop structure forms base-stacking interactions with the zipped stem

We next sought to further characterize the alternative stem-loop structure we discovered by identifying stabilizing interactions in $\boldsymbol{Z^L}$ that are not present in $\boldsymbol{Z^S}$. Unzipping along the alternative pathway resulted in a greater disruption of interactions than unzipping along the native pathway (Figure 3a). This greater disruption in stabilizing interactions suggested that the stem itself is more stable and resistant to unzipping in the alternative pathway than in the native pathway. We stress that this increased stability refers to the stem alone; when considering the stabilizing interactions contributed by both the stem and the loop, the natively folded $\boldsymbol{Z^S}$ is more stable overall than the alternatively folded $\boldsymbol{Z^L}$. Nonetheless, this greater stability of the stem in $\boldsymbol{Z^L}$ relative to $\boldsymbol{Z^S}$ is counterintuitive based on the widely described role of the native UUCG loop structure in preferentially stabilizing the zipped stem-loop [13, 26, 28, 29].

Since stem unzipping is not expected to lead to any changes in the loop structure itself, the increased stability of the zipped stem in $\boldsymbol{Z^L}$ must arise from the alternative loop structure forming more interactions with the stem than the native loop structure. Previous computational studies guided our identification of these interactions [25]. In the native loop structure, G_{L4} in the UUCG loop (Figure 1a) adopts the *syn*-configuration to form the *trans*-G-U wobble base pair. An *anti*-configuration of G_{L4} would inhibit the formation of this *trans*-G-U wobble interaction, but instead allow G_{L4} to stack against the adjacent G_{CBP} (G in the closing base pair; Figure 1a) [47]. We hypothesize that it is this stacking interaction in $\boldsymbol{Z^L}$ that stabilizes the zipped stem and needs to be overcome during unzipping in the alternative pathway (Figure 3a). Intriguingly, the stronger stabilizing interactions that the stem makes with the alternative loop structure *versus* the native loop structure, suggests that the role of the native loop structure in stem-loop folding goes beyond the formation of these interactions.

Formation of the closing base-pair determines the rate of stem zipping

To explore the role of the native loop structure in stem-loop folding, we sought to identify the transition state ($\boldsymbol{TS}^{\ddagger}$) RNA conformations that comprise the barriers to the zipping and unzipping transitions. We used our temperature-dependent measurements of zipping to describe these barriers in terms of changes in stabilizing interactions (*i.e.*, activation enthalpy, $\Delta H^{\ddagger}$s) and in conformational flexibility (*i.e.*, activation entropy, $\Delta S^{\ddagger}$s) during the rate-determining steps of these transitions [48]. For zipping transitions in both pathways, the negative $\Delta S^{\ddagger}$s indicate that the $\boldsymbol{TS}^{\ddagger}$s are significantly more constrained than the corresponding unzipped conformations (Figures 3b and 3c). Thus, these barriers are dominated by the challenge of constraining the stem in order to zip. In contrast, the $\boldsymbol{TS}^{\ddagger}$ s are almost equally as constrained as the corresponding zipped conformations, but contain less stabilizing interactions (Figures 3b and 3c). Specifically, these $\boldsymbol{TS}^{\ddagger}$s both possess half of the stabilizing interactions that are required for the stem to fully zip

(Figures 3a and 3b, and Table S6). Given that zipping of the stem in our construct involves formation of only two base-pairs, this result suggests that one, and only one, of these base-pairs is already formed in the $\boldsymbol{TS}^{\ddagger}$s, and it is the formation of this base-pair, associated with a significant loss in flexibility of the RNA, that determines the rate of stem zipping in both pathways.

Comparison of the two pathways revealed that formation of the first base-pairing interaction leads to greater stabilization (*i.e.*, a larger $\Delta H^{\ddagger}$) in $\boldsymbol{TS}_{\boldsymbol{L}}^{\ddagger}$ than in $\boldsymbol{TS}_{\boldsymbol{S}}^{\ddagger}$ (Figure 3b). We attribute this increased stabilization to the additional base-stacking interaction formed between G_{L4} and G_{CBP} in $\boldsymbol{TS}_{\boldsymbol{L}}^{\ddagger}$. Since this stacking interaction is likely to be formed only when G_{CBP} is base-paired, it is specifically the closing base pair that is formed in $\boldsymbol{TS}_{\boldsymbol{L}}^{\ddagger}$. Given the overall similarity between zipping in the alternative and native pathways, we expect $\boldsymbol{TS}_{\boldsymbol{S}}^{\ddagger}$ to also contain just the closing base pair. The formation of the second base pair also leads to greater stabilization between $\boldsymbol{TS}_{\boldsymbol{L}}^{\ddagger}$ and $\boldsymbol{Z}^{\boldsymbol{L}}$, than between $\boldsymbol{TS}_{\boldsymbol{S}}^{\ddagger}$ and $\boldsymbol{Z}^{\boldsymbol{S}}$ (Figure 3b), as would be expected when the stacking network in $\boldsymbol{TS}_{\boldsymbol{L}}^{\ddagger}$ subsequently expands to the outer base pairs in $\boldsymbol{Z}^{\boldsymbol{L}}$.

Our results thus reveal that it is the loss of flexibility associated with formation of the closing base-pair of the stem that serves as the barrier for zipping. On the other hand, the barrier for unzipping is almost exclusively due to the disruption of the outer base-pair (Figure 3a), and therefore, is not associated with any significant increase in flexibility (Figure 3c). The small differences in entropy that we observe between the $\boldsymbol{Z}$ conformations and the $\boldsymbol{TS}^{\ddagger}$s (*i.e.* the $\Delta S^{\ddagger}$s for unzipping) likely arise from changes in the ordering of the solvent when the outer base pair forms [49] rather than from changes in the flexibility of stem-loop itself.

The native loop structure facilitates zipping by increasing the proximity between the closing bases

The results described above suggest that any RNA structure that increases the proximity between the two closing bases would facilitate zipping by reducing the change in flexibility that is needed to form the stem. This is the case for $\boldsymbol{U}^{\boldsymbol{S}}$, where formation of the first base-pair is associated with

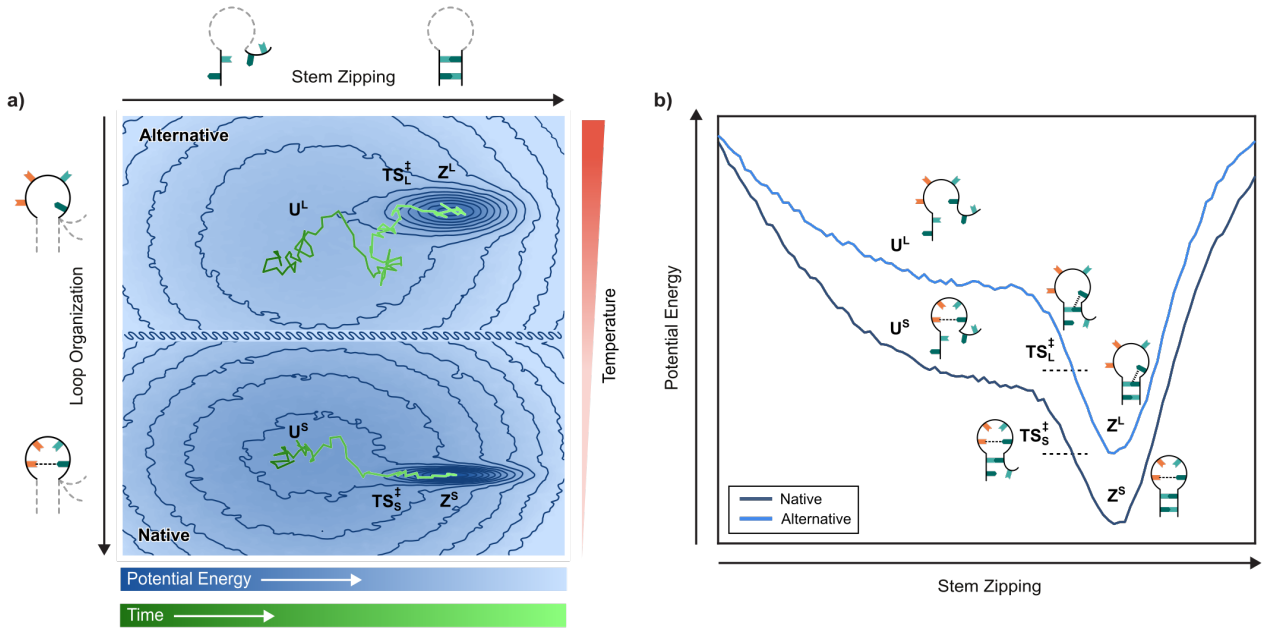

Figure 4. The physical basis of UUCG stem-loop stability. (a) *The hypothetical hierarchical potential energy landscape for the folding of the stem-loop along the two reaction coordinates of loop organization and stem zipping. Each point in the landscape represents the potential energy of a specific stem-loop conformation projected onto these coordinates. The border between the native and alternative loop structure basin is shown as discontinuous to represent the lack of information about the barriers between the two in our analysis of the data. Two hypothetical tracks representing stem-loop folding in both basins are shown, a short track capturing the fast guided search for* $\boldsymbol{Z^S}$ *in the narrow native loop structure basin, and a more difussive track representing the slower search for* $\boldsymbol{Z^L}$ *in the wider alternative loop structure basin. As temperature is increased, the alternative loop structure basin is preferred.* ***(b)*** *A hypothetical representation of the one-dimensional potential energy along the stem zipping reaction coordinate for the native and alternative pathways. The positions of* $\boldsymbol{U^L}, \boldsymbol{Z^L}$*, and* $\boldsymbol{TS_L^{\ddagger}}$ *in the alternative pathway, and* $\boldsymbol{U^S}, \boldsymbol{Z^S}$*, and* $\boldsymbol{TS_S^{\ddagger}}$ *in the native pathway, are shown for both* ***(a)*** *and* ***(b)****.*

a smaller loss of flexibility than for the fully unfolded $\boldsymbol{U^L}$ (Figure 3c). The pre-organization of the native loop structure in $\boldsymbol{U^S}$ thus leads to a faster rate of zipping from $\boldsymbol{U^S}$ (Figure 1d). Counterintuitively, in the alternative pathway, the stacking interaction between G_{L4} and G_{CBP} leads to a further loss of flexibility that needs to be overcome to form the stem, resulting in a slower rate of zipping despite the greater stabilization that results from forming this interaction.

A powerful way to visualize how the native loop structure facilitates stem-loop folding is through a schematic of the potential energy landscape for the folding mechanism of the UUCG stem-loop (Figure 4a). In this hierarchical landscape, two large basins represent the ensembles of conformations in which the UUCG tetraloop adopts the native or alternative loop structure. These basins each contain one low-potential-energy well corresponding to $\boldsymbol{Z^L}$ or $\boldsymbol{Z^S}$, while regions

of the basins outside each well correspond to $\boldsymbol{U^L}$ or $\boldsymbol{U^S}$. The depths of the basins and wells represent the relative stabilities due to the interactions present in the corresponding conformations, while their widths represent their relative flexibilities. The natively folded $\boldsymbol{Z^S}$, which has more stabilizing interactions and is more constrained than $\boldsymbol{Z^L}$, corresponds to a deeper and narrower well. Since our existing data analysis tools cannot identify the transitions between the two basins in our experimental data (*i.e.*, the transitions between the slow and fast phases of zipping and unzipping in our current trajectories), we represent the barrier between the two basins as a discontinuity.

The process of stem-loop folding from the fully unfolded $\boldsymbol{U^L}$ to natively folded $\boldsymbol{Z^S}$ is the conformational search on this landscape from the shallow and wide non-native loop basin into the deep and narrow $\boldsymbol{Z^S}$ well. The UUCG stem-loop has evolved to pay the cost associated with the loss of flexibility for this transition in two steps, using a set of hierarchical loop and stem dynamics. In the first step, part of this cost is paid by organizing the loop into the native structure, forming a metastable intermediate, $\boldsymbol{U^S}$, that nucleates the process of stem-loop folding, and shifts to the deeper and, in particular, narrower native loop basin. This constrains the conformational search, allowing transitions to the $\boldsymbol{Z^S}$ well from $\boldsymbol{U^S}$ to be faster (as shown by the shorter track in Figure 4a). The rate-limiting step for zipping of the stem *via* this pathway is the second step, requiring loss of the remaining flexibility in the RNA to reach $\boldsymbol{TS_S^{\ddagger}}$, which contains only the closing base-pair and is situated at the entrance to the $\boldsymbol{Z^S}$ well. The downhill pathway from $\boldsymbol{U^L}$, through $\boldsymbol{U^S}$**,** to $\boldsymbol{Z^S}$ allows the loss of flexibility at each of the two steps to be coupled to a gain in stabilizing interactions (Figure 4b).

In summary, the ability of the UUCG sequence to form the native loop structure in the absence of a stem facilitates folding of the stem-loop into $\boldsymbol{Z^S}$ in two ways. First, it creates a hierarchical separation between loop organization and stem zipping such that the entropic cost of stem-loop folding can be paid over two steps, resulting in a guided conformational search for the

natively folded structure of the stem-loop. Second, by pre-organizing the loop structure to nucleate folding in the first step, formation of the closing base-pair in $\boldsymbol{TS_S^{\ddagger}}$ is made more probable, cooperatively facilitating the second step of stem formation. Taken together, this describes the mechanism through which pre-organization of the tetraloop significantly reduces the conformational space of the stem-loop and expedites the folding process.

The alternative zipped conformation resists unfolding through compensatory changes in flexibility

In addition to the guided-search mechanism of folding along the native pathway described above, we see that the alternative folding pathway also contributes significantly to the observed stability of the UUCG stem-loop. This is specifically observed under conditions, such as higher temperatures, where the alternative zipping pathway becomes progressively more dominant (Tables S4 and S5). This effect can also be visualized using the potential energy landscape for UUCG stem-loop folding (Figure 4a). The basin for the alternative zipping pathway is very similar to the native one described above, with a $\boldsymbol{Z^L}$ well, and a transition state, $\boldsymbol{TS_L^{\ddagger}}$, at the entrance of the well (Figure 4a). Importantly, however, the $\boldsymbol{Z^L}$ well is wider than the $\boldsymbol{Z^S}$ well. When faced with destabilizing perturbations like higher temperatures, the conformational search for this well from the fully unfolded $\boldsymbol{U^L}$ is easier than for the narrower $\boldsymbol{Z^S}$ well. This is represented in the potential energy landscape by the more diffusive track in the alternative pathway basin—representing more thermal fluctuations of the RNA—still being able to find and zip into the wider $\boldsymbol{Z^L}$ well (Figure 4a). Secondly, the wider $\boldsymbol{Z^L}$ well also reflects how the stem-loop can adopt a larger variation of more flexible loop structures while maintaining a zipped stem, which is crucial for remaining folded at higher temperatures. Finally, due to the additional base stacking interaction between the stem and the loop, $\boldsymbol{Z^L}$ possesses a higher barrier to exit than $\boldsymbol{Z^S}$, and once in the well, transitioning out is less probable (Figure 4b). Altogether, this shows how the stem-loop can conditionally favor

more flexible RNA conformations by shifting its ensemble to the more flexible $\boldsymbol{Z^L}$ in order to resist unfolding at higher temperatures.

$\boldsymbol{Z^L}$ is clearly a kinetic trap in the folding process of the native UUCG stem-loop structure. However, the crucial role of this conformation in the stability of the folded stem-loop structure as a whole leads us to suggest that $\boldsymbol{Z^L}$ does not represent merely a misfolded conformation, but is instead a previously unidentified, functionally relevant alternative structure of this RNA. While beyond the scope of this work, additional functional roles for this alternative UUCG loop structure need to be further explored. We note, for instance, that the alternative loop structure in $\boldsymbol{Z^L}$ (see above, Figure 3a) may position the UUCG stem-loop to make tertiary contacts with other RNAs or proteins by freeing the bases in the loop. This previously uncharacterized conformation may thus have unexplored biological roles beyond stabilizing the stem-loop by resisting unfolding.

Discussion

We have elucidated the complete conformational ensemble of a fundamental RNA structural motif, the UUCG stem-loop, using a high-time-resolution, single-molecule experimental approach uniquely suited for studying the dynamics of such biopolymers. We characterize, in molecular detail, the competing pathways, metastable intermediates, and kinetic traps that collaborate to give rise to the exceptional thermodynamic stability of this RNA stem-loop. Our findings reveal that the UUCG stem-loop uses a hierarchical conformational search process to guide folding into the fully folded native structure. We further show how the specific ordering of this hierarchy into the two steps of loop structure organization and stem zipping facilitates the efficient collapse of the flexible RNA into the constrained native folded state (Figure 4a) in the native zipping pathway—an entirely entropic effect. We further discover how this RNA stem-loop resists unfolding when challenged with destabilizing perturbations by shifting its equilibrium to a more flexible alternatively folded conformation (Figure 4a), a conformation that may also have

additional, hitherto uncharacterized biological roles. Our results elucidate how these dual mechanisms of constrained conformational search and compensatory changes to flexibility collaborate to give rise to the exceptional thermodynamic stability of the UUCG stem-loop.

The folding mechanism we have elucidated here has clear implications for our views of the biological roles of UNCG stem-loops (where N denotes any nucleotide). UNCG stem-loops, including UUCG stem-loops, are typically considered to only serve as ultra-stable structural scaffolds [15]. This perspective has predominately been informed by static structures of the fully folded, native structure of the UUCG stem-loop, which exhibits only a limited set of tertiary interactions [13, 26, 28, 29]. However, our results show that this native structure is not the sole folded structure that can be populated by this stem-loop and, indeed, is the less populated structure under certain conditions. In contrast to the native loop structure, the alternative loop structure we observe here is expected to allow for greater tertiary interactions. Our results thus suggest that the UUCG stem-loop may play a biological role in interacting with other biomolecules through this alternative structure. Our hypothesis is supported by a recent study analyzing existing structures of RNA stem-loops, which found that UNCG tetraloops can populate multiple conformations referred to as 'non-native' when bound to other biomolecules [50]. In particular, the ability of the UUCG stem-loop to shift between the native and alternative loop structures in response to perturbations sets up an enticing model for the regulation of stem-loop interactions with other biomolecules. In this model, perturbations to the UUCG stem-loop, whether through changing conditions or the allosteric effect of the binding of ligands or other biomolecules, can shift the equilibrium between the two loop structures, tuning the ability of the stem-loop to interact with partners. This, of course, is a hypothesis which needs to be further investigated in light of our findings.

Finally, the complete folding mechanism of the UUCG stem-loop that we reveal here (Figure 4) is in striking contrast to our current understanding, which attributes the exceptional thermodynamic stability of this RNA motif solely to the interactions within the unique native UUCG

loop structure [13, 28-30]. While beyond the scope of this work, it will be worthwhile to determine how this folding mechanism maps onto other thermodynamically stable RNA motifs, including GNRA stem-loops (where N denotes any nucleotide and R any purine nucleotide), which are thought to have evolved a strategy for achieving their exceptional thermodynamic stability that is independent and distinct from that of the UNCG stem-loops [51]. It is worth noting, in this context, that the strategies for thermodynamic stability we describe here do not appear to be specific to stem-loops—or indeed, to RNA—and thus, we expect to see them in the folding mechanisms of a wide range of biomolecules. In particular, the strategy we have discovered here of resisting destabilizing perturbations through compensatory changes in the flexibility of the biomolecule represents an exciting addition to our understanding of biomolecular stability, the implications of which must be explored for broader classes of structured biopolymers.

Materials and Methods

Design of the smFET experiments

The single-walled carbon nanotube (SWCNT)-based smFET experimental platform, including details describing the theoretical basis for the function of an smFET device, practical aspects of how the current is measured, and how the temperature is controlled, has been previously described [34, 38, 52]. Briefly, an smFET device provides a measure of current through the SWCNT as a function of time. This current signal is extremely sensitive to the local charge environment near the surface of the SWCNT where the biomolecule is tethered. Because conformational changes, including folding and unfolding, of a biomolecule such as an RNA alters the electrostatic charge distribution near the surface of the SWCNT, the current *vs.* time trajectory (hereafter referred to as the current trajectory) recorded by an smFET device very sensitively reports on these conformational changes.

The smFET experiments reported in this work utilized a chemically synthesized RNA stem-loop construct (Integrated DNA Technologies) that we refer to here as h-SL-bio (Table S1). h-SL-bio consisted of a stem containing two G-C base-pairs and a loop composed of a four-nucleotide UUCG 'tetraloop' sequence. The 5' end of h-SL-bio was modified with a hydrazide functional group that allowed us to tether the construct to the surface of the SWCNT of our smFET devices. The 3' end of the h-SL-bio was further functionalized with a biotin for potential downstream validation of tethering with additional markers. Fortunately, tethering of h-SL-bio to the surface of SWCNT could be directly validated using the current trajectories recorded by the smFET device (see below and Figures S1 and S2) and therefore downstream validation of tethering using the biotin functional group was unnecessary. To confirm that the 3' biotin moiety does not alter the overall thermodynamic stability of the UUCG stem-loop, we conducted additional thermal melting experiments in which we compared the h-SL-bio construct to the same chemically synthesized RNA construct (Integrated DNA Technologies) that did not contain the biotin modification (a construct that we refer to as h-SL, Table S1). The presence of the biotin did

not perturb the thermodynamic properties of the stem-loop, demonstrating that it did not alter the folding dynamics of the stem-loop (Figure S3 and Table S2).

A single molecule of the h-SL-bio construct was tethered to the SWCNT surface using an electrically coupled diazonium reaction [34, 35] (Figure S1a). Briefly, a source-drain current *vs.* liquid-gate voltage (I_{sd}-V_{lg}) curve was obtained from the smFET device prior to running the diazonium reaction by applying a constant source-drain voltage (V_{sd}) and a sweeping V_{lg} from −0.5 V to +0.5 V through the platinum electrode of the smFET device in Recording Buffer (10 mM sodium phosphate (Na_2HPO_4)/sodium dihydrogen phosphate (NaH_2PO_4), pH = 7.0). Before the introduction of 4-formylbenzene diazonium hexafluorophosphate (FBDP) to the devices, the device was placed at a V_{lg} of −0.5 V to prevent spurious diazonium reactions. The device was subsequently incubated in 300 μM FBDP dissolved in Recording Buffer to initiate the reaction. All devices were then slowly ramped towards a positive V_{lg} until a clear downward transition in current was observed, indicating the coupling of FBDP to the SWCNT surface (Figure S1b). Following the FBDP reaction, all devices were rinsed with 1 mL of Recording Buffer at 50 °C to remove any residual FBDP salts, and a new I_{sd}-V_{lg} curve was obtained with the same conditions as before the diazonium reaction. The comparison between the I_{sd}-V_{lg} curves shows a clear reduction in conductance before and after FBDP exposure (Figure S1c). To tether each stem-loop construct, the hydrazide functional group at the 5' end of the stem-loop construct was coupled to the aldehyde functional group of the SWCNT-tethered FDBP *via* a hydrazide-aldehyde coupling reaction. To achieve this, the device was incubated in 100 μM of the stem-loop construct in Reaction Buffer (100 mM sodium acetate (NaOAc), pH=4.5) overnight (for ~20 hours) followed by a 2-hour reduction in Reduction Buffer (10 mM $NaHPO_4$/$Na_2H_2PO_4$,, pH = 7.0 and 50 mM sodium cyanoborohydride ($NaCNBH_3$)). The device was then rinsed twice with 500 μL of Recording Buffer at 50 °C to remove any non-specifically bound RNA. The smFET experimental data in this work derives from this single RNA molecule tethered to the smFET device.

The smFET experiments were performed in Recording Buffer. Current trajectories were recorded at varying temperatures (35, 40, 45, and 50 °C) for 5 minutes at a V_{lg} of 0.1 V and a V_{sd} of 25 mV. To achieve the desired temperature, the microfluidic chamber was incubated in the heating chamber of the smFET instrument for 15 min at the specified temperature before data collection was initiated. The smFET device was sampled at a time resolution of 40 μs using a 5.3 kHz anti-aliasing filter.

Thermal melting experiments

To validate that our single SWCNT-tethered RNA molecule recapitulated the ensemble thermodynamic stability of the UUCG stem-loop, we performed thermal melting experiments of the h-SL-bio construct by monitoring the circular dichroism (CD) of the RNA at 280 nm as a function of temperature over a temperature window from 5 °C to 95 °C using a Chirascan V100 CD spectrophotometer. For this thermal melting experiment, a 10 μM sample of the h-SL-bio construct was prepared in Recording Buffer. Prior to the thermal melting experiment, the stem-loop construct was folded by pre-heating to 95 °C and subsequent cooling to 5 °C at a ramp rate of 2 °C min^{-1}. CD measurements were recorded at every 0.2 °C increment with a ramp rate of 1 °C min^{-1} (Figure S3). Using a two-state model comprised of unimolecular transitions between single $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations [20], the equilibrium changes in enthalpy ($\Delta H_{eq}^{(2)}$) and entropy ($\Delta S_{eq}^{(2)}$) of folding for the stem-loop, where the superscript of (2) is used to denote the two-state model described above, was calculated using the following set of equations:

$$Y = (m_U T + b_U) + [(m_Z T + b_Z) - (m_U T + b_U)] \frac{K_{eq}^{(2)}}{1 + K_{eq}^{(2)}} ,$$

$$K_{eq}^{(2)} = e^{\frac{\Delta H_{eq}^{(2)}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)} \text{ , and}$$

$$\Delta S_{eq}^{(2)} = \frac{\Delta H_{eq}^{(2)}}{T_m},$$

where Y is the observed CD signal, m_U and m_Z are the baseline slopes for the $\boldsymbol{U}$ and $\boldsymbol{Z}$ states, b_U and b_Z are the corresponding baseline intercepts, $K_{eq}^{(2)}$ is the two-state equilibrium constant for the zipping reaction, R is the universal gas constant (1.987 cal K^{-1} mol^{-1}), T is the absolute temperature of the measurement, and T_m is the absolute 'melting' temperature at which the corresponding free energy $\Delta G_{eq}^{(2)}$ (given by $\Delta G_{eq}^{(2)} = \Delta H_{eq}^{(2)} - \mathrm{T}\Delta S_{eq}^{(2)}$) is zero. The above equations were fit to data collected for two independent replicates of the thermal melting experiment using a nonlinear curve fitting algorithm implemented with Matlab to yield the values of $\Delta H_{eq}^{(2)}$ and $\Delta S_{eq}^{(2)}$ for the stem-loop (Table S2). These values were subsequently used to calculate the $\Delta G_{eq}^{(2)}$ for the UUCG stem-loop at the temperatures for the smFET experiments (Figure S4).

An additional set of thermal melting experiments were performed for the h-SL construct in a manner identical to the h-SL-bio construct described above (Figure S3). The data from two replicates of this thermal melting experiment was similarly used to estimate the $\Delta H_{eq}^{(2)}$ and $\Delta S_{eq}^{(2)}$ for the h-SL construct (Table S2).

Identification of current states within current trajectories

The current trajectories for all of the experimental temperatures (35, 40, 45, and 50 °C; Figure S2) were first individually background corrected using an iterative detection algorithm (as previously described in [34, 36, 53]) to correct for the slowly drifting current baseline inherent to smFET devices [34, 36, 38]. The current trajectories at each temperature were subsequently divided into 200-ms time intervals, each containing 5,000 data points. Each interval was normalized such that the lowest current value for the interval corresponded to a normalized current signal of 0 a.u., and the highest current value corresponded to 1 a.u. This normalization further corrected for long-time-scale changes in the signal due to any remaining background current drift in the smFET device [36]. Intervals where excessive local background current drift and/or noise in the smFET device obscured transitions in the current trajectory resulting in an

inability to clearly distinguish the two current states were discarded. In total, such intervals represented less than 3-10% of the total intervals collected at each temperature point.

At all temperatures other than 35 °C, the two current states present in the current trajectories could be identified and characterized using Gaussian mixture models (GMMs) inferred by employing a variational Bayesian expectation-maximization algorithm [54] implemented in the single-molecule data analysis platform, tMAVEN [55]. These GMMs yielded the means and standard deviations of the two observed states (Figure S2), which were subsequently used to generate idealized trajectories from the experimental data (based on the responsibilities [54] of the individual data points for each Gaussian distribution). The high-current state was centered around a normalized current value of ~0.7 a.u. and the low-current state was centered around a normalized current value of ~0.3 a.u. For the 35 °C dataset, the relative proportion of the low-current state was so small that the signal for this state could not be reliably described as a Gaussian distribution using a GMM. In this case, a threshold of 0.4 a.u. (which corresponded to a normalized current value three standard deviations below the mean of the high-current state) was utilized to differentiate between the two states. Based on a previous smFET study of RNA stem-loops, the high-current state was assigned to zipped ($\boldsymbol{Z}$) conformations of the stem-loop, and the low-current state was assigned to unzipped ($\boldsymbol{U}$) conformations [36].

Inference of kinetic rates from current trajectories

Based on the idealized trajectories calculated as described above, dwell times in both $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations (defined as the contiguous time period the molecule spent in a $\boldsymbol{U}$ or $\boldsymbol{Z}$ conformation before transitioning to a $\boldsymbol{Z}$ or $\boldsymbol{U}$ conformation, respectively) were calculated. These dwell times were then used to generate the corresponding survival probability distributions, $F(t)$, defined as the fraction of dwells that were greater than a specific t, for each conformation (Figures S5 and S6). Each survival probability distribution was fit to a single-exponential decay of the form

$$y(t) = Ae^{-kt} ,$$

where k represents the rate constant and A is the pre-exponential term and a double-exponential decay of the form

$$y(t) = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} ,$$

where k_1 and k_2 represent the rate constants for the two kinetic phases, and A_1 and A_2 are the respective pre-exponential terms for the two phases that also give their relative proportions. Subsequently, residuals of the fits,

$$r(t) = F(t) - y(t) ,$$

were also calculated. These residuals showed that the single-exponential fits missed the initial fast decay phase of the survival probability distribution for both conformations at all temperatures (Figures S5 and S6). Thus, the double-exponential decays better explained the survival distributions. The double-exponential decays each yielded a slow phase (k_{slow} and A_{slow}) that arises from a long-lived RNA conformation ($\boldsymbol{U^L}$ or $\boldsymbol{Z^L}$) and a fast phase (k_{fast} and A_{fast}) that arises from a short-lived RNA conformation ($\boldsymbol{U^S}$ or $\boldsymbol{Z^S}$). The kinetic parameter values we obtained from the double-exponential fits at all temperatures for the zipping and unzipping transitions are compiled in Tables S4 and S5, respectively.

Calculation of activation enthalpies and entropies using transition state theory

Each k was converted to the corresponding activation free energy ($\Delta G^{\ddagger}$) (Figure S7a-b) using the relation specified by transition state theory [48]

$$k = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} ,$$

where T is the absolute temperature, k_B is the Boltzmann constant, h is the Planck constant, R is the universal gas constant, and κ is a correction factor that, in our calculations, is set to 1 [48]. For each kinetic phase in both the zipping and the unzipping transitions, the corresponding $\Delta G^{\ddagger}$s

over the experimental range of temperatures were used to estimate the $\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$s using a linear fit to

$$\Delta G^{\ddagger}(T) = \Delta H^{\ddagger} - T\Delta S^{\ddagger}.$$

The $\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$s were assumed to be temperature-independent in the relatively small temperature range of the reported experiments, which agreed with the observed linear dependence of $\Delta G^{\ddagger}$ on temperature (Figure S7a-b). The calculated $\Delta G^{\ddagger}$s for both phases at all temperatures, and the fitted $\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$s, are compiled for the zipping and unzipping transitions in Table S6.

Calculation of the equilibrium enthalpies and entropies between the kinetic phases

For both $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations, the relative proportions of the slow and fast kinetic phases were given by their respective pre-exponential terms (A_{slow} and A_{fast}) in the double-exponential fits (Tables S4 and S5). Since these phases arose from the individual populations of the long-lived ($\boldsymbol{U^L}$ or $\boldsymbol{Z^L}$) and short-lived ($\boldsymbol{U^S}$ or $\boldsymbol{Z^S}$) conformations, the ratios between these two terms ($\frac{A_{slow}}{A_{fast}}$) were interpreted as the ratios between the proportions of these long-lived and short-lived RNA conformations. This allowed us to calculate the equilibrium constants (K_{eq}) between the two $\boldsymbol{U}$ conformations (and similarly, the two $\boldsymbol{Z}$ conformations) as

$$K_{eq} = \frac{A_{slow}}{A_{fast}}.$$

Subsequently, these K_{eq}s, calculated over varying temperatures, were converted to the corresponding free energy differences, ΔG_{eq}, at these temperatures as

$$\Delta G_{eq} = -RT \ln K_{eq},$$

where T is the absolute temperature and R is the universal gas constant. Similar to the activation parameters, these ΔG_{eq}s (Figure S7c-d) were fit to the linear equation

$$\Delta G_{eq}(T) = \Delta H_{eq} - T\Delta S_{eq},$$

where ΔH_{eq} and ΔS_{eq} are the corresponding equilibrium enthalpic and entropic differences, that are also assumed to be temperature independent in the relatively small experimental range of temperatures studied here. The calculated ΔG_{eq}s at all temperatures, and the fitted ΔH_{eq}s and ΔS_{eq}s, are compiled in Table S7.

Acknowledgements

We thank Dr. Hashim al-Hashimi (Columbia University) and Dr. Jeffrey Kieft (New York Structural Biology Center) for valuable comments on the manuscript. We also thank Dr. Sarah Dubnik and other past and present members of the Gonzalez, Nuckolls, and Shephard labs for their assistance, comments, and advice. This work was carried out in part in the Clean Room, Electron Microscopy, and Shared Materials Characterization labs of Columbia Nano Initiative (CNI) Shared Lab Facilities and the Precision Biomolecular Characterization Facility (PBCF) at Columbia University.

Funding

C.N. thanks Sheldon and Dorothea Buckler for their generous support. This research was supported by the National Science Foundation (CHE 2004016) and the National Institutes of Health (GM107417 and GM153724). Essential instrumentation in the PBCF was made possible by funding from the National Institutes of Health (S10OD025102).

Authors Contributions

S.S.J., K.K.R., D.G.L., C.N., and R.L.G. designed the research; S.S.J. and D.G.L. fabricated the smFET devices; S.S.J. and D.G.L. performed the smFET experiments and collected the data; S.S.J. and K.K.R. analyzed the data; S.S.J., K.K.R., D.G.L., C.N., and R.L.G. wrote the manuscript; all authors approved the final version of the manuscript.

Competing Interests

D.G.L. and K.L.S. have a financial interest in Quicksilver Biosciences, Inc., which is commercializing smFET technology for molecular diagnostic applications. The other authors declare no competing financial interest.

Data and materials availability

The open-source Python code for tMAVEN, which was used for the single-molecule data analysis in this work, is freely available *via* a Git repository (https://github.com/GonzalezBiophysicsLab/tmaven). The current trajectories will be made freely available via Zenodo repository at the time of publication.

References

1. Tinoco, I. and C. Bustamante, *How RNA folds.* Journal of Molecular Biology, 1999. **293**(2): p. 271-281.
2. Dill, K.A., et al., *The protein folding problem.* Annu Rev Biophys, 2008. **37**: p. 289-316.
3. Mustoe, A.M., C.L. Brooks, and H.M. Al-Hashimi, *Hierarchy of RNA functional dynamics.* Annu Rev Biochem, 2014. **83**: p. 441-66.
4. Ganser, L.R., et al., *The roles of structural dynamics in the cellular functions of RNAs.* Nat Rev Mol Cell Biol, 2019. **20**(8): p. 474-489.
5. Herschlag, D., S. Bonilla, and N. Bisaria, *The Story of RNA Folding, as Told in Epochs.* Cold Spring Harb Perspect Biol, 2018. **10**(10).
6. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The Energy Landscapes and Motions of Proteins.* Science, 1991. **254**(5038): p. 1598-1603.
7. Vicens, Q. and J.S. Kieft, *Thoughts on how to think (and talk) about RNA structure.* Proceedings of the National Academy of Sciences, 2022. **119**(17): p. e2112677119.
8. Fraser, J.S. and M.A. Murcko, *Structure is beauty, but not always truth.* Cell, 2024. **187**(3): p. 517-520.
9. Wei, G., et al., *Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell.* Chemical Reviews, 2016. **116**(11): p. 6516-6551.
10. Bloomfield, V.A., D.M. Crothers, and I. Tinoco, *Nucleic Acids: Structure, Properties, and Functions*. G - Reference,Information and Interdisciplinary Subjects Series. 2000: University Science Books.
11. Brion, P. and E. Westhof, *Hierarchy and dynamics of RNA folding.* Annu Rev Biophys Biomol Struct, 1997. **26**: p. 113-37.
12. Antao, V.P., S.Y. Lai, and I. Tinoco, *A thermodynamic study of unusually stable RNA and DNA hairpins.* Nucleic Acids Research, 1991. **19**(21): p. 5901-5905.
13. Varani, G., C. Cheong, and I. Tinoco, Jr., *Structure of an unusually stable RNA hairpin.* Biochemistry, 1991. **30**(13): p. 3280-9.
14. Woese, C.R., S. Winker, and R.R. Gutell, *Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops".* Proceedings of the National Academy of Sciences, 1990. **87**(21): p. 8467-8471.
15. Hall, K.B., *Mighty tiny.* RNA, 2015. **21**(4): p. 630-1.
16. Yesselman, J.D., et al., *Computational design of three-dimensional RNA structure and function.* Nat Nanotechnol, 2019. **14**(9): p. 866-873.
17. Bevilacqua, P.C. and J.M. Blose, *Structures, kinetics, thermodynamics, and biological functions of RNA hairpins.* Annu Rev Phys Chem, 2008. **59**: p. 79-103.
18. Yang, X., et al., *Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping.* Nat Struct Biol, 2001. **8**(11): p. 968-73.
19. Sarkar, K., D.A. Nguyen, and M. Gruebele, *Loop and stem dynamics during RNA hairpin folding and unfolding.* Rna, 2010. **16**(12): p. 2427-34.

20. Proctor, D.J., et al., *Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate.* Biochemistry, 2004. **43**(44): p. 14004-14.
21. Sarkar, K., et al., *Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe.* Biophys J, 2009. **97**(5): p. 1418-27.
22. Bargmann, C.I., *Comparative chemosensation from receptors to ecology.* Nature, 2006. **444**(7117): p. 295-301.
23. Mrazikova, K., et al., *UUCG RNA Tetraloop as a Formidable Force-Field Challenge for MD Simulations.* J Chem Theory Comput, 2020. **16**(12): p. 7601-7617.
24. Sorin, E.J., et al., *RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop.* J Mol Biol, 2002. **317**(4): p. 493-506.
25. Chen, A.A. and A.E. García, *High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations.* Proc Natl Acad Sci U S A, 2013. **110**(42): p. 16820-5.
26. Nozinovic, S., et al., *High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA.* Nucleic Acids Research, 2010. **38**(2): p. 683-694.
27. Chakraborty, D., R. Collepardo-Guevara, and D.J. Wales, *Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins.* J Am Chem Soc, 2014. **136**(52): p. 18052-61.
28. Ennifar, E., et al., *The crystal structure of UUCG tetraloop.* Journal of Molecular Biology, 2000. **304**(1): p. 35-42.
29. Cheong, C., G. Varani, and I. Tinoco, Jr., *Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC.* Nature, 1990. **346**(6285): p. 680-2.
30. Nozinovic, S., et al., *High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA.* Nucleic Acids Res, 2010. **38**(2): p. 683-94.
31. Ma, H., et al., *Exploring the energy landscape of a small RNA hairpin.* J Am Chem Soc, 2006. **128**(5): p. 1523-30.
32. Sorin, E.J., et al., *Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations.* Biophys J, 2003. **85**(2): p. 790-803.
33. Childs-Disney, J.L., et al., *Targeting RNA structures with small molecules.* Nat Rev Drug Discov, 2022. **21**(10): p. 736-762.
34. Lee, Y., et al., *Electrically Controllable Single-Point Covalent Functionalization of Spin-Cast Carbon-Nanotube Field-Effect Transistor Arrays.* ACS Nano, 2018. **12**(10): p. 9922-9930.
35. Wilson, H., et al., *Electrical Monitoring of sp3 Defect Formation in Individual Carbon Nanotubes.* The Journal of Physical Chemistry C, 2016. **120**(3): p. 1971-1976.
36. Jang, S.S., et al., *Characterizing the Conformational Free-Energy Landscape of RNA Stem-Loops Using Single-Molecule Field-Effect Transistors.* Journal of the American Chemical Society, 2023. **145**(1): p. 402-412.
37. Sorgenfrei, S., et al., *Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor.* Nat Nanotechnol, 2011. **6**(2): p. 126-32.
38. Vernick, S., et al., *Electrostatic melting in a single-molecule field-effect transistor with applications in genomic identification.* Nat Commun, 2017. **8**: p. 15450.
39. Choi, Y., et al., *Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering.* Nano Lett, 2013. **13**(2): p. 625-31.
40. Wallace, M.I., et al., *Non-Arrhenius kinetics for the loop closure of a DNA hairpin.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5584-9.
41. Ansari, A., S.V. Kuznetsov, and Y. Shen, *Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins.* Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7771-6.
42. Munoz, V., et al., *Folding dynamics and mechanism of beta-hairpin formation.* Nature, 1997. **390**(6656): p. 196-9.
43. Howe, C.P., et al., *Sequence-Dependent Melting and Refolding Dynamics of RNA UNCG Tetraloops Using Temperature-Jump/Drop Infrared Spectroscopy.* J Phys Chem B, 2023. **127**(7): p. 1586-1597.
44. Howe, C.P., et al., *Measuring RNA UNCG Tetraloop Refolding Dynamics Using Temperature-Jump/Drop Infrared Spectroscopy.* The Journal of Physical Chemistry Letters, 2022. **13**(39): p. 9171-9176.
45. Borkar, A.N., et al., *Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop.* Phys Chem Chem Phys, 2017. **19**(4): p. 2797-2804.

46. Bottaro, S., et al., *Integrating NMR and simulations reveals motions in the UUCG tetraloop.* Nucleic Acids Res, 2020. **48**(11): p. 5839-5848.
47. Chen, A.A. and A.E. García, *High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations.* Proceedings of the National Academy of Sciences, 2013. **110**(42): p. 16820-16825.
48. Fersht, A., *Structure and Mechanism in Protein Science*. Series in Structural Biology. Vol. Volume 9. 2017: WORLD SCIENTIFIC. 656.
49. Dunitz, J.D., *The Entropic Cost of Bound Water in Crystals and Biomolecules.* Science, 1994. **264**(5159): p. 670-670.
50. Bottaro, S. and K. Lindorff-Larsen, *Mapping the Universe of RNA Tetraloop Folds.* Biophys J, 2017. **113**(2): p. 257-267.
51. Heus, H.A. and A. Pardi, *Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops.* Science, 1991. **253**(5016): p. 191-4.
52. Bouilly, D., et al., *Single-Molecule Reaction Chemistry in Patterned Nanowells.* Nano Lett, 2016. **16**(7): p. 4679-85.
53. Plesa, C. and C. Dekker, *Data analysis methods for solid-state nanopores.* Nanotechnology, 2015. **26**(8): p. 084003.
54. Bishop, C.M., *Pattern recognition and machine learning*. 2006: New York : Springer, [2006] ©2006.
55. Verma, A.R., et al., *Increasing the accuracy of single-molecule data analysis using tMAVEN.* Biophys J, 2024.

SUPPORTING INFORMATION

RNA adapts its flexibility to efficiently fold and resist unfolding

Sukjin S. Jang[1,†,§], Korak Kumar Ray[1,‡,§], David G. Lynall[2,§], Kenneth L. Shepard[2], Colin Nuckolls[1,*], Ruben L. Gonzalez, Jr.[1,*]

[1] *Department of Chemistry, Columbia University, New York, NY 10027 USA*

[2] *Department of Electrical Engineering, Columbia University, New York, NY 10027 USA*

[†] *Current Address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115 USA*

[‡] *Current Address: MRC-Laboratory of Medical Sciences, London W12 0HS, UK*

[§] These authors contributed equally to this work.

*To whom correspondence may be addressed: Colin Nuckolls, Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, Tel: (212) 854-6289, Email: cn37@columbia.edu.; Ruben L. Gonzalez, Jr., Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, Tel: (212) 854-1096, Email: rlg2118@columbia.edu

Conductance trajectories arise from the zipping dynamics of a single stem-loop

To validate that the observed conductance trajectories in our experiments arise from the zipping dynamics of a single, surface-tethered stem-loop RNA, we first determined the ensemble, two-state, differences in free energies, $\Delta G_{eq}^{(2)}$s, for zipping of the UUCG stem-loop, where the superscript denotes our assumption of the two-state model, using thermal melting experiments (Figure S3; see Materials and Methods). To determine the analogous $\Delta G_{eq}^{(2)}$s for zipping of the UUCG stem-loop using our smFET data, we applied a Gaussian mixture model (GMM)-based kinetic model [1] that identified two current states in our conductance trajectories. For the lowest temperature point of 35 °C, due to the very small population of the low-current state, a threshold was used, instead of a GMM, to distinguish between the two states (see Materials and Methods). Assuming that each state corresponded to a single RNA conformation, we were able to assign to the low- and high-current states to the unzipped $\boldsymbol{U}$ and the zipped $\boldsymbol{Z}$ conformations of the stem-loop (Figures 1 and S2). We note that the application of the GMM for the higher temperature points was made possible by an improved signal-to-noise ratio (SNR) of ~4–5 in this work relative to the SNR of ~3-4 that we have observed in our previous smFET studies of RNA stem-loops [2]. This improvement could be attributed to several changes in the experimental design of the present work *versus* our previous work, including the use of a lower ionic strength buffer, resulting in reduced screening of the electrostatic charges on the RNA [3, 4]; the use of an RNA stem-loop construct where the conformational changes induced by zipping and unzipping were more proximal to the surface of the single-walled carbon nanotube (SWCNT); and the direct, covalent functionalization of the RNA stem-loop construct to the surface of the SWCNT [5-7] (Figure S1). These improvements over our previous work increased the change in current through the SWCNT during the conformational rearrangements of the RNA molecule, resulting in the higher SNR observed in the present work.

Using our GMM- and threshold-based kinetic models, we calculated the fractional occupancies of the $\boldsymbol{Z}$ and $\boldsymbol{U}$ conformations (f_Z and f_U, respectively) across the temperatures investigated here (Figure S2). In agreement with a previous study [2], f_U was observed to increase with increasing temperature, which is in line with the expectation that the unzipped $\boldsymbol{U}$ conformation should become more populated with increasing temperature (Table S3). These fractional occupancies were subsequently used to calculate the two-state equilibrium constants, $K_{eq}^{(2)}s$, for the zipping reaction using

$$K_{eq}^{(2)} = \frac{f_Z}{f_U},$$

and the corresponding two-state equilibrium free energies of zipping using

$$\Delta G_{eq}^{(2)} = -RT \ln K_{eq}^{(2)},$$

where R denotes the universal gas constant and T is the absolute temperature of the measurement. The $\Delta G_{eq}^{(2)}$s for the smFET experiments were in striking agreement with those from the ensemble thermal melting experiments (Figure S4a and Table S3), validating that the current trajectories obtained from our smFET experiments report on the same zipping reaction that the ensemble thermal melting experiments report on.

Next, using the $K_{eq}^{(2)}$s estimated from the thermal melting experiments, we calculated the theoretical fractional occupancies of the $\boldsymbol{U}$ conformation for a given number of surface-tethered stem-loops over the range of experimental temperatures investigated here (Figure S4b). In the case of a single RNA stem-loop tethered to a single smFET device, f_U was directly calculated from $K_{eq}^{(2)}$ as

$$f_U = \frac{1}{1 + K_{eq}^{(2)}}.$$

For the case of multiple RNA stem-loops tethered to a single smFET device, we assumed that the $\boldsymbol{U}$ conformation corresponded to the scenario where at least one stem-loop was unzipped and

the $\boldsymbol{Z}$ conformation corresponded to the case where all stem-loops were zipped. Given these assumptions, the expected fractional occupancy of the $\boldsymbol{U}$ conformation for this case was given by

$$f_U = 1 - f_Z^N,$$

where N denotes the number of tethered stem-loops and f_Z is the fractional occupancy of a single RNA stem-loop at a given temperature, which can be calculated from the ensemble $K_{eq}^{(2)}$ using

$$f_Z = \frac{K_{eq}^{(2)}}{1 + K_{eq}^{(2)}}.$$

Comparing the f_U values obtained from the GMM-based kinetic modeling employed to analyze for our experimentally obtained current trajectories with the f_U values that would be expected for various different numbers of surface-tethered RNA stem-loops allowed direct determination of the number of RNA stem-loops that generated a particular, experimentally obtained current trajectory. The results of this comparison showed that the f_U values we experimentally obtained overwhelmingly agree with the f_U values expected for the case in which a single RNA stem-loop is tethered to the surface of the SWCNT, confirming that our experimental conductance trajectories arose from a single, surface-tethered RNA molecule (Figure S4b). We note that the analysis described here may be used in the future as a general framework for directly determining the number of surface-tethered biomolecules on any smFET device.

Identification of the kinetic scheme between observed RNA conformations

To determine the kinetic scheme governing the zipping dynamics of the UUCG stem-loop, we made use of the fact that equilibrium changes in enthalpy and entropy (ΔH_{eq} and ΔS_{eq}, respectively) are thermodynamic state variables. Thus, for any two given states, these quantities should depend only on the identities of the states and not on the paths used to calculate them. We considered the two $\boldsymbol{U}$ conformations ($\boldsymbol{U^L}$ and $\boldsymbol{U^S}$) and two different paths connecting them, the 'direct' path A and the 'indirect' path B (Figures 2 and S7). For path A, $\Delta H_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ and

$\Delta S_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ between these conformations, where $\boldsymbol{U^S}$ represents the initial conformation and $\boldsymbol{U^L}$ the final conformation, were calculated directly from the proportion of the slow and fast kinetic phases that corresponded to these $\boldsymbol{U}$ conformations (Figure S8c-d and Table S7; see Materials and Methods).

The indirect path B involves the two $\boldsymbol{Z}$ conformations ($\boldsymbol{Z^L}$ and $\boldsymbol{Z^S}$) (Figures 2 and S7). In this path, $\Delta H_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ and $\Delta S_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ can be expressed as differences between the $\boldsymbol{U}$ conformations and the $\boldsymbol{Z}$ conformations, and between the $\boldsymbol{Z}$ conformations themselves [8]. As above, $\Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S})$ and $\Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S})$ can be directly calculated based on the proportion of the slow and fast kinetic phases that corresponded to these $\boldsymbol{Z}$ conformations (Figure S8c-d and Table S7; see Materials and Methods). However, the ΔH_{eq} s (and ΔS_{eq} s) between the $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations cannot be calculated in a similar manner, because our dwell-time analysis (Figures S5 and S6) only yielded the rates of transitions out of a particular conformation and not the rates between two conformations [9]. Instead, these rates yielded the activation parameters ($\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$) for transitions out of these conformations (Figure S8a-b and Table S6; see Materials and Methods). These $\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$ represented the corresponding changes between the $\boldsymbol{U}$,and between the $\boldsymbol{Z}$, conformations and the transition states ($\boldsymbol{TS}^{\ddagger}$s) for the transitions exiting these conformations [10]. To calculate the corresponding ΔH_{eq}s and ΔS_{eq}s, we need to first specify how these $\boldsymbol{TS}^{\ddagger}$s fit into the kinetic scheme.

For kinetic scheme 1 (Figures 2a and S7c), we assume that the $\boldsymbol{TS}^{\ddagger}$ exiting out of $\boldsymbol{U^S}$ ($\boldsymbol{TS_1^{\ddagger}}$) corresponds to that out of $\boldsymbol{Z^S}$. We can then estimate the ΔH_{eq} and ΔS_{eq} between them as

$$\Delta H_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^S}) = \Delta H^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}) - \Delta H^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^S}) \text{ , and}$$

$$\Delta S_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^S}) = \Delta S^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}) - \Delta S^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^S}).$$

In this case, the other $\boldsymbol{TS}^{\ddagger}$ ($\boldsymbol{TS_2^{\ddagger}}$) out of $\boldsymbol{U^L}$ would correspond to that out of $\boldsymbol{Z^L}$, resulting in

$$\Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^L}) = \Delta H^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}) - \Delta H^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^L}) \text{ , and}$$

$$\Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^L}) = \Delta S^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}) - \Delta S^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^L}).$$

Thus, the 'indirect' $\Delta H_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ and $\Delta S_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S})$ can be written as

$$\Delta H_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S}) = \Delta H_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^S}) + \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^L})$$
$$= \Delta H^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}\right) - \Delta H^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^S}\right) + \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta H^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}\right) + \Delta H^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^L}\right)$$

and

$$\Delta S_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S}) = \Delta S_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^S}) + \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^L})$$
$$= \Delta S^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}\right) - \Delta S^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^S}\right) + \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta S^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}\right) + \Delta S^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^L}\right).$$

In this scheme, by placing the $\boldsymbol{TS^{\ddagger}}$s in the above manner, we have assumed that transitions only occur between $\boldsymbol{U^S}$ and $\boldsymbol{Z^S}$, and between $\boldsymbol{U^L}$ and $\boldsymbol{Z^L}$ (Figures 2a and S7c).

Alternatively, in kinetic scheme 2 (Figures 2b and S7d), we assume that the $\boldsymbol{TS^{\ddagger}}$ exiting out of $\boldsymbol{U^S}$ ($\boldsymbol{TS_1^{\ddagger}}$) corresponds to that out of $\boldsymbol{Z^L}$, resulting in

$$\Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^S}) = \Delta H^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}) - \Delta H^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^L}) \text{ and}$$

$$\Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^S}) = \Delta S^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}) - \Delta S^{\ddagger}(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^L}) \ .$$

In this case, the other $\boldsymbol{TS^{\ddagger}}$ out of $\boldsymbol{U^L}$ ($\boldsymbol{TS_2^{\ddagger}}$) would correspond to that out of $\boldsymbol{Z^S}$, resulting in

$$\Delta H_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^L}) = \Delta H^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}) - \Delta H^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^S}) \text{ and}$$

$$\Delta S_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^L}) = \Delta S^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}) - \Delta S^{\ddagger}(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^S}) \ .$$

Thus, in this case, the 'indirect' ΔH_{eq} and ΔS_{eq} between $\boldsymbol{U^L}$ and $\boldsymbol{U^S}$ can be written as

$$\Delta H_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S}) = \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^S}) - \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta H_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^L})$$
$$= \Delta H^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}\right) - \Delta H^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^L}\right) - \Delta H_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta H^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}\right) + \Delta H^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^S}\right)$$

and

$$\Delta S_{eq}(\boldsymbol{U^L}, \boldsymbol{U^S}) = \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{U^S}) - \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta S_{eq}(\boldsymbol{Z^S}, \boldsymbol{U^L})$$
$$= \Delta S^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{U^S}\right) - \Delta S^{\ddagger}\left(\boldsymbol{TS_1^{\ddagger}}, \boldsymbol{Z^L}\right) - \Delta S_{eq}(\boldsymbol{Z^L}, \boldsymbol{Z^S}) - \Delta S^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{U^L}\right) + \Delta S^{\ddagger}\left(\boldsymbol{TS_2^{\ddagger}}, \boldsymbol{Z^S}\right).$$

In this scheme, with the above placement of the $\boldsymbol{TS^{\ddagger}}$s, we have assumed that transitions only occur between $\boldsymbol{U^S}$ and $\boldsymbol{Z^L}$, and between $\boldsymbol{U^L}$ and $\boldsymbol{Z^S}$ (Figures 2b and S7d).

While the two schemes differ only in how the $\boldsymbol{TS}^{\ddagger}$s, and thus the corresponding activation parameters, are arranged, we see that they lead to drastically different expressions for the indirect ΔH_{eq} and ΔS_{eq} between $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$. By analyzing our experimentally obtained data in this way, we show that the zipping dynamics of the UUCG stem-loop are predominantly governed by kinetic scheme 1 (Figures 2 and S7c-d, right-most panels).

References

1. Verma, A.R., et al., *Increasing the accuracy of single-molecule data analysis using tMAVEN.* Biophys J, 2024.
2. Jang, S.S., et al., *Characterizing the Conformational Free-Energy Landscape of RNA Stem-Loops Using Single-Molecule Field-Effect Transistors.* Journal of the American Chemical Society, 2023. **145**(1): p. 402-412.
3. Sorgenfrei, S., et al., *Debye screening in single-molecule carbon nanotube field-effect sensors.* Nano Lett, 2011. **11**(9): p. 3739-43.
4. Choi, Y., et al., *Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering.* Nano Lett, 2013. **13**(2): p. 625-31.
5. Wilson, H., et al., *Electrical Monitoring of sp3 Defect Formation in Individual Carbon Nanotubes.* The Journal of Physical Chemistry C, 2016. **120**(3): p. 1971-1976.
6. Vernick, S., et al., *Electrostatic melting in a single-molecule field-effect transistor with applications in genomic identification.* Nat Commun, 2017. **8**: p. 15450.
7. Lee, Y., et al., *Electrically Controllable Single-Point Covalent Functionalization of Spin-Cast Carbon-Nanotube Field-Effect Transistor Arrays.* ACS Nano, 2018. **12**(10): p. 9922-9930.
8. Zumdahl, S.S. and S.A. Zumdahl, *Chemistry*. 2006: Cengage Learning.
9. Kinz-Thompson, C.D., N.A. Bailey, and R.L. Gonzalez, Jr., *Precisely and Accurately Inferring Single-Molecule Rate Constants.* Methods Enzymol, 2016. **581**: p. 187-225.
10. Fersht, A., *Structure and Mechanism in Protein Science*. Series in Structural Biology. Vol. Volume 9. 2017: WORLD SCIENTIFIC. 656.

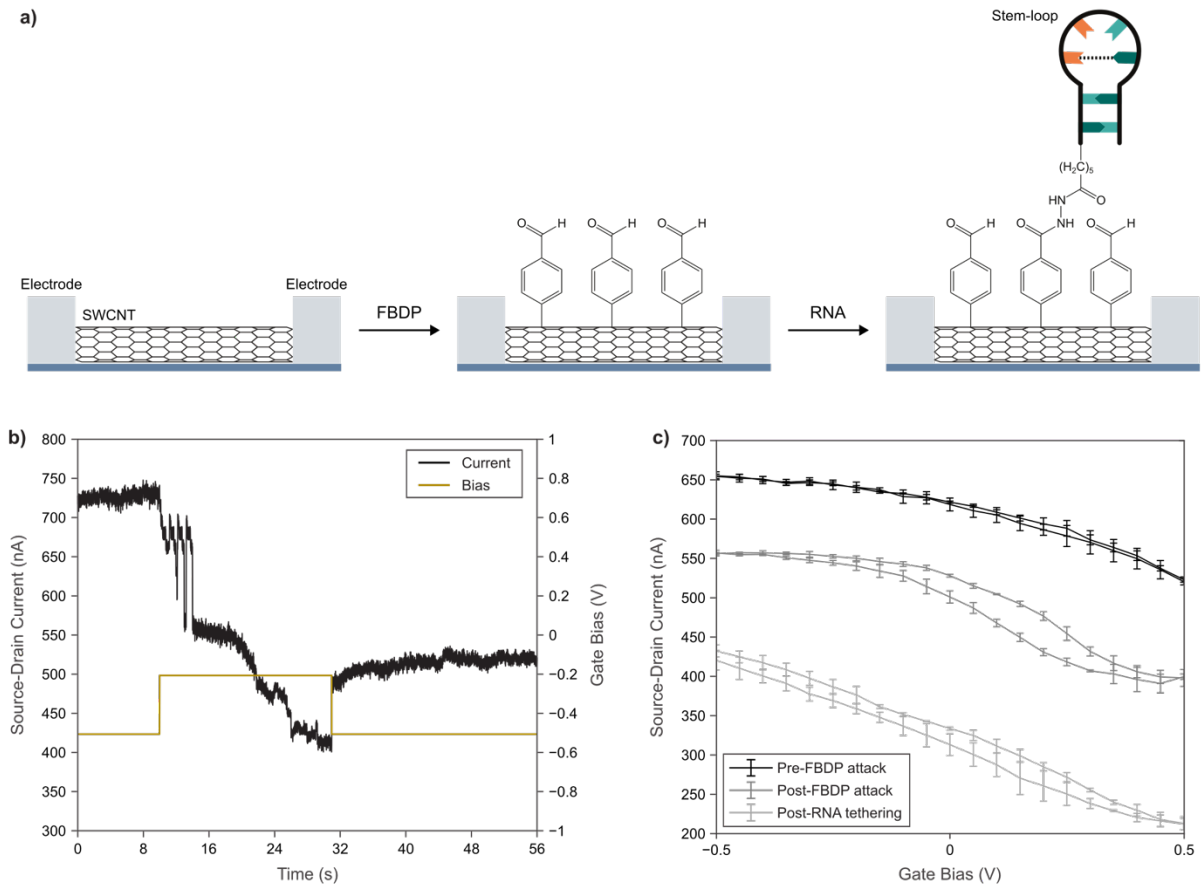

Figure S1. Covalent tethering of an RNA molecule to the surface of a SWCNT using an electrically controlled diazonium coupling reaction. (a) Schematic of the covalent tethering approach, showing first the tethering of FBDP onto the SWCNT surface through an electrically controlled diazonium reaction, followed by an RNA tethering step through a aldehyde-hydrazide coupling reaction. **(b)** I_{sd}-t (black) and V_{lg}-t (brown) measurements of an smFET device during the FBDP attack, showing multiple, clearly distinguished drops in current indicative of FBDP tethering when V_{lg} is ramped to −0.2 V from −0.5 V. **(c)** Average I_{sd}-V_{lg} curves for the same smFET device before FBDP attack, after FBDP attack, and after RNA tethering, measured in Recording Buffer. Error bars denote the standard deviation of three measurements.

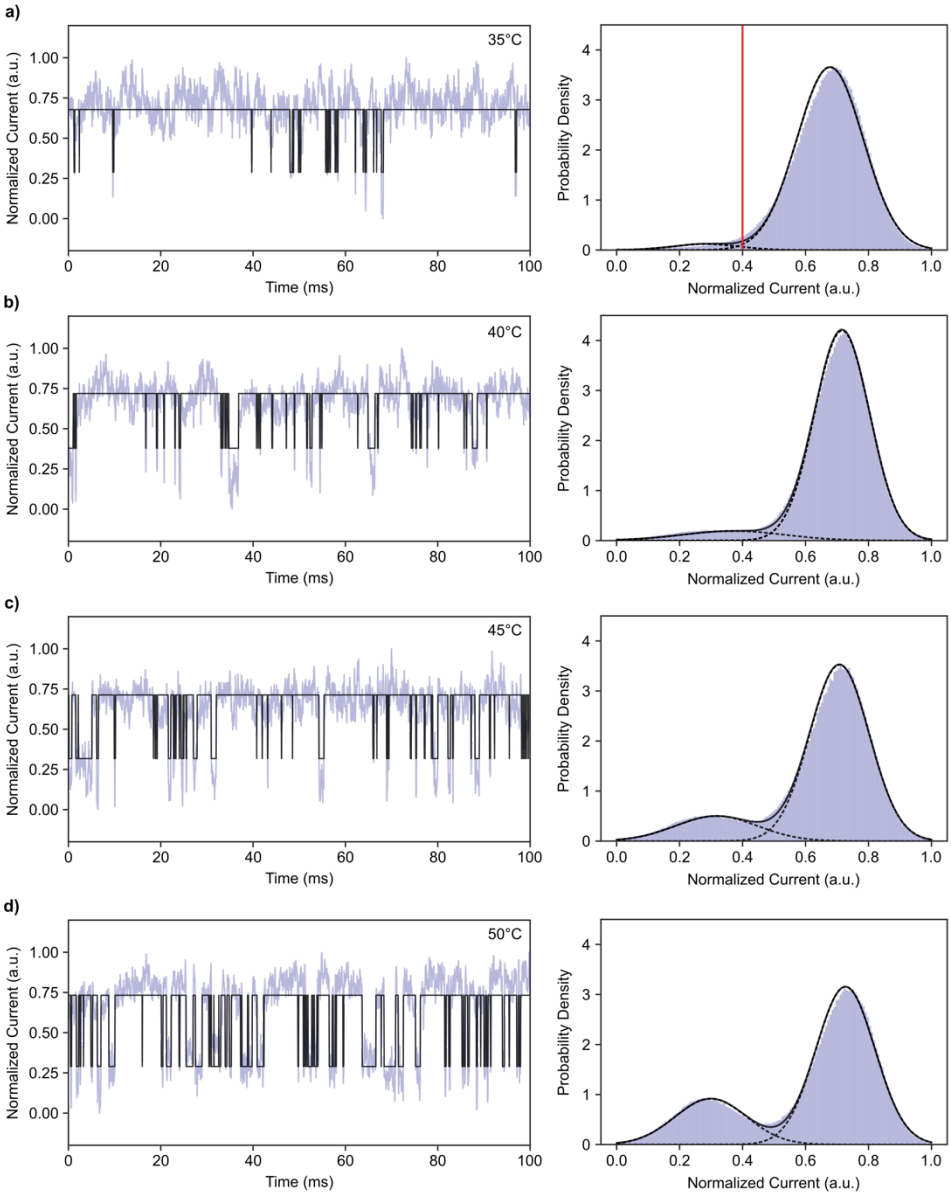

Figure S2. Temperature dependence of current trajectories recorded using the same, single UUCG stem-loop tethered onto the SWCNT surface of the same, single smFET device. smFET datasets obtained at **(a)** 35 °C, **(b)** 40 °C, **(c)** 45 °C, and **(d)** 50 °C showing (*left*) a representative 100-ms interval of the current trajectory (violet) overlaid with the idealized trajectory (black) and (*right*) a histogram of the current values across an entire 60 s trajectory (violet) overlaid with the inferred Gaussian density for each current state (dashed black curves) and the sum of all states (solid black curve) for the dataset at each temperature. The kinetic model that yielded these Gaussian distributions at 40 °C, 45 °C, and 50 °C were GMMs that were inferred from the data. For the 35 °C dataset, current values were first separated by a threshold value at 0.4 a.u. (red) and the means and standard deviations of the resulting populations yielded the individual Gaussian distributions for each state.

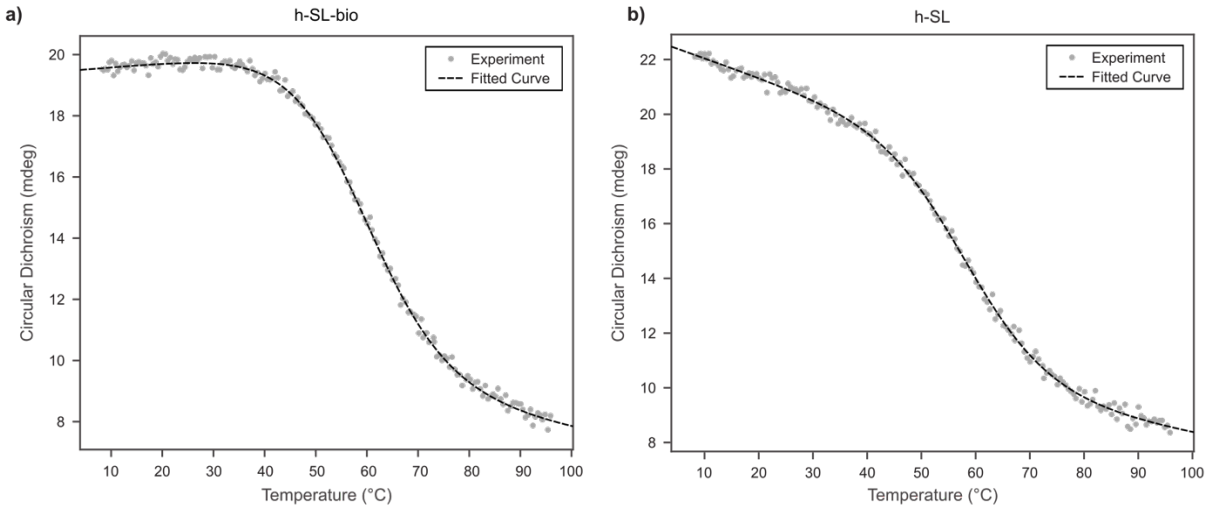

Figure S3. Thermal melting of the h-SL-bio and h-SL RNA stem-loop constructs. Thermal melting curves showing the CD measured at 280 nm (gray circles) for the **(a)** h-SL-bio and **(b)** h-SL constructs as a function of temperature, along with the fit (black curve) to a two-state model of unimolecular transitions between the $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations.

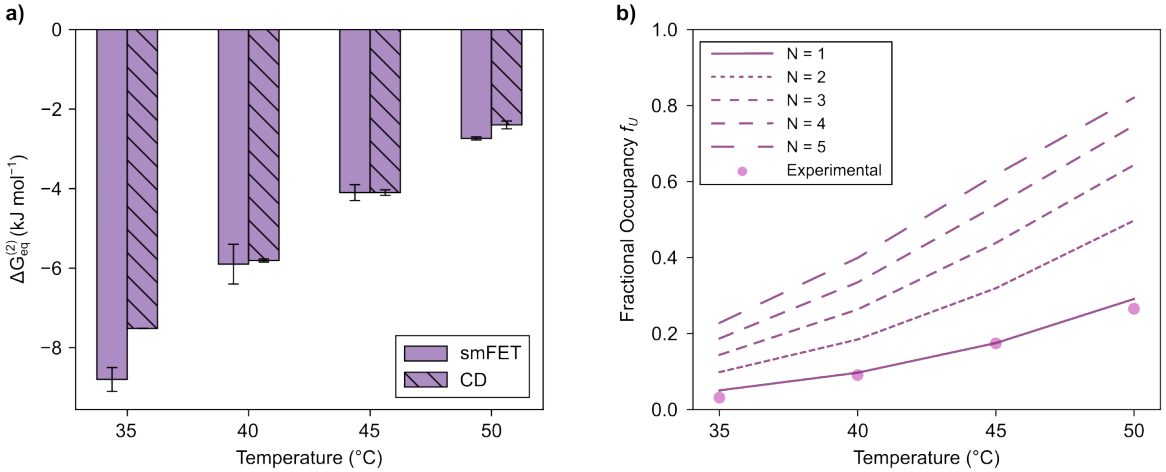

Figure S4. Confirmation that conductance trajectories report on the zipping dynamics of single, surface-tethered RNA stem-loops. (a) Comparisons of the $\Delta G_{eq}^{(2)}$s of zipping for the smFET (solid) and CD thermal melting (hatched) experiments. Error bars for the smFET experiments represent the standard deviations of the $\Delta G_{eq}^{(2)}$s calculated from three separate 20-s windows for the current trajectories at each temperature. Error bars for the CD thermal melting experiments represent the standard deviations of the $\Delta G_{eq}^{(2)}$s calculated from two independent experiments. **(b)** The expected fractional occupancies of the $\boldsymbol{U}$ conformation (f_U) at each temperature for a single tethered RNA stem-loop and various multiple tethered RNA stem-loop (lines) along with the f_U calculated from experimental smFET data (circles).

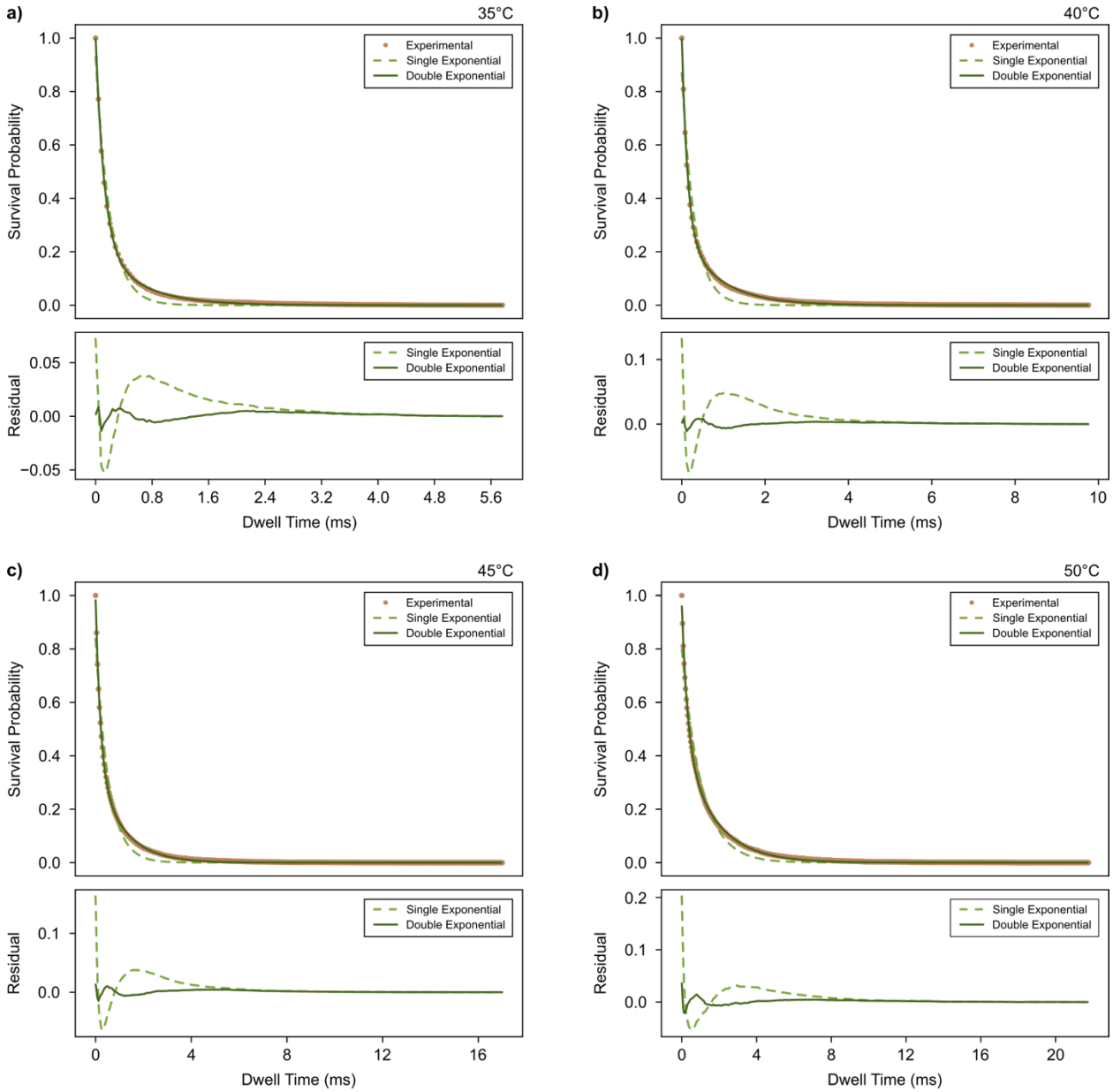

Figure S5. Kinetics of zipping. (*top*) Survival probabilities for the dwell-time distributions of the $\boldsymbol{U}$ conformation (circle) along with the single-exponential (dashed) and double-exponential (solid) fits and (*bottom*) the residuals of the single-exponential (dashed) and double-exponential (solid) fits for smFET datasets collected at **(a)** 35 °C, **(b)** 40 °C, **(c)** 45 °C, and **(d)** 50 °C.

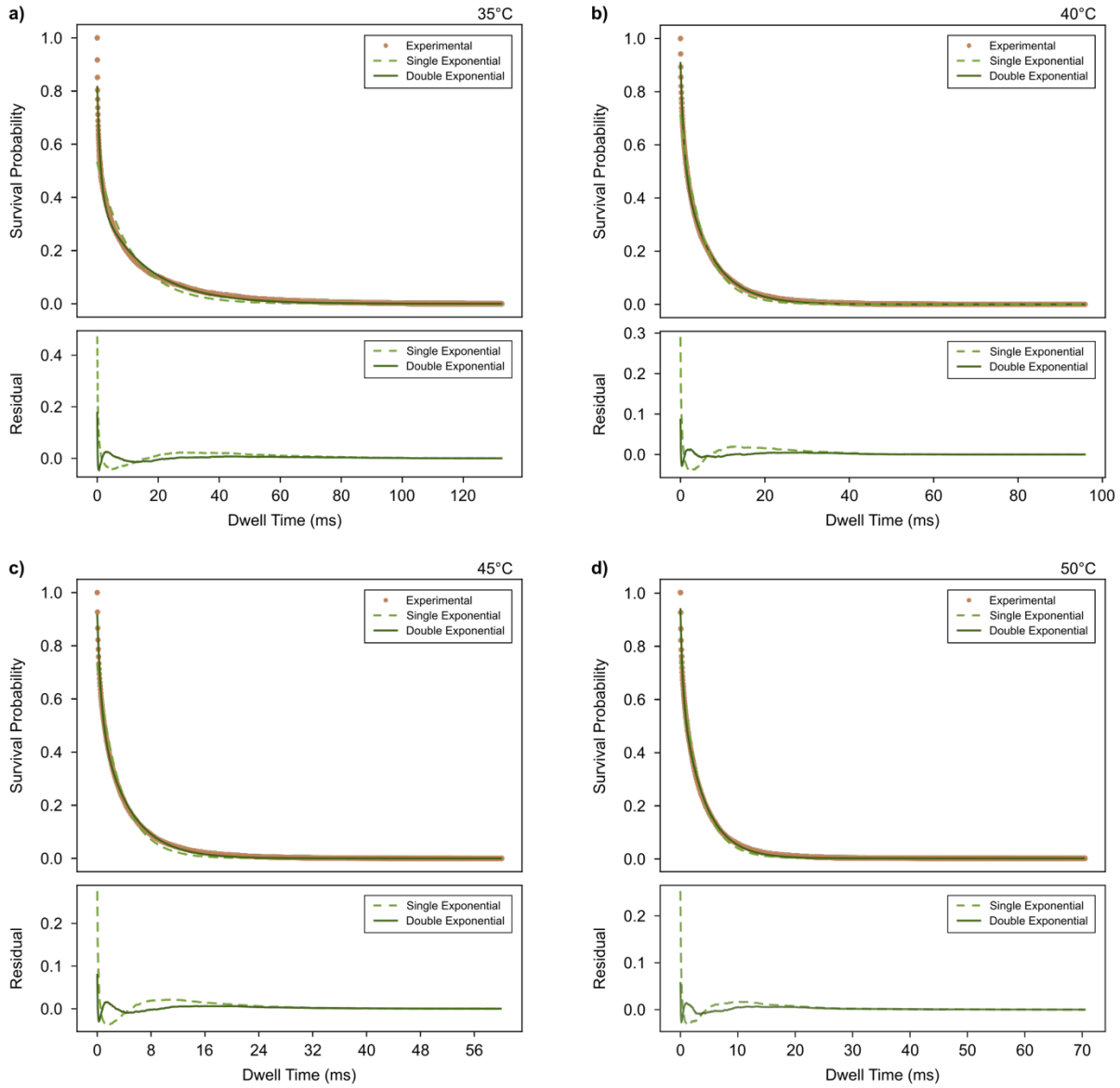

Figure S6. Kinetics of unzipping. (*top*) Survival probabilities for the dwell-time distributions of the ***Z*** conformation (circle) along with the single-exponential (dashed) and double-exponential (solid) fits and (*bottom*) the residuals of the single-exponential (dashed) and double-exponential (solid) fits for smFET datasets collected at **(a)** 35 °C, **(b)** 40 °C, **(c)** 45 °C, and **(d)** 50 °C.

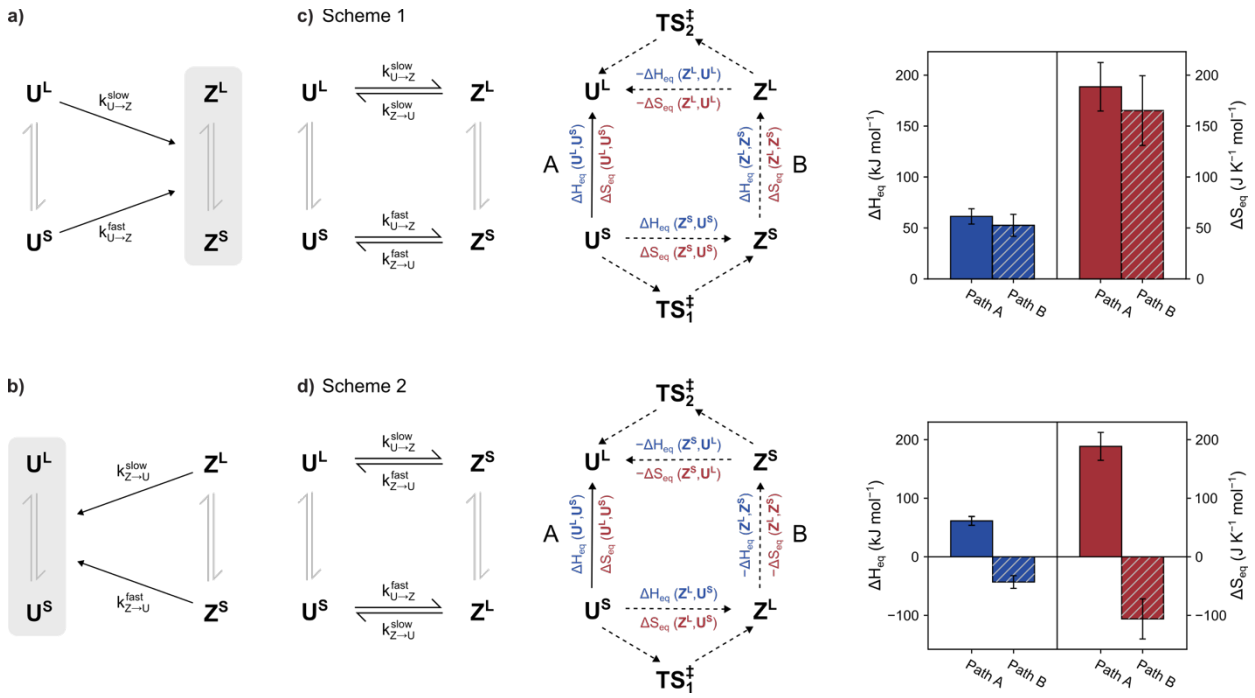

Figure S7. Potential kinetic schemes governing the zipping dynamics of the UUCG RNA stem-loop. The observed slow and fast kinetic phases, and the corresponding conformations that they transition from, for **(a)** zipping and **(b)** unzipping. **(c)** Potential kinetic scheme 1, showing *(left)* the predominant transitions between $\boldsymbol{U^S}$ and $\boldsymbol{Z^S}$ and between $\boldsymbol{U^L}$ *and* $\boldsymbol{Z^L}$, *(center)* the corresponding thermodynamic cycle depicting the direct path A (solid arrow) and the indirect path B (dashed arrows) between $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$, and *(right)* the calculated ΔH_{eq}s (blue) and ΔS_{eq}s (red) for these paths. The unobserved transitions between the $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations are shown as grey equilibrium arrows. **(d)** Potential kinetic scheme 2, showing *(left)* the predominant transitions between $\boldsymbol{U^S}$ *and* $\boldsymbol{Z^L}$ and between $\boldsymbol{U^L}$ and $\boldsymbol{Z^S}$, *(center)* the corresponding thermodynamic cycle depicting the direct path A (solid arrow) and the indirect path B (dashed arrows) between $\boldsymbol{U^S}$ and $\boldsymbol{U^L}$, and *(right)* the calculated ΔH_{eq}s (blue) and ΔS_{eq}s (red) for these paths. The unobserved transitions between the $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations are shown as grey equilibrium arrows.

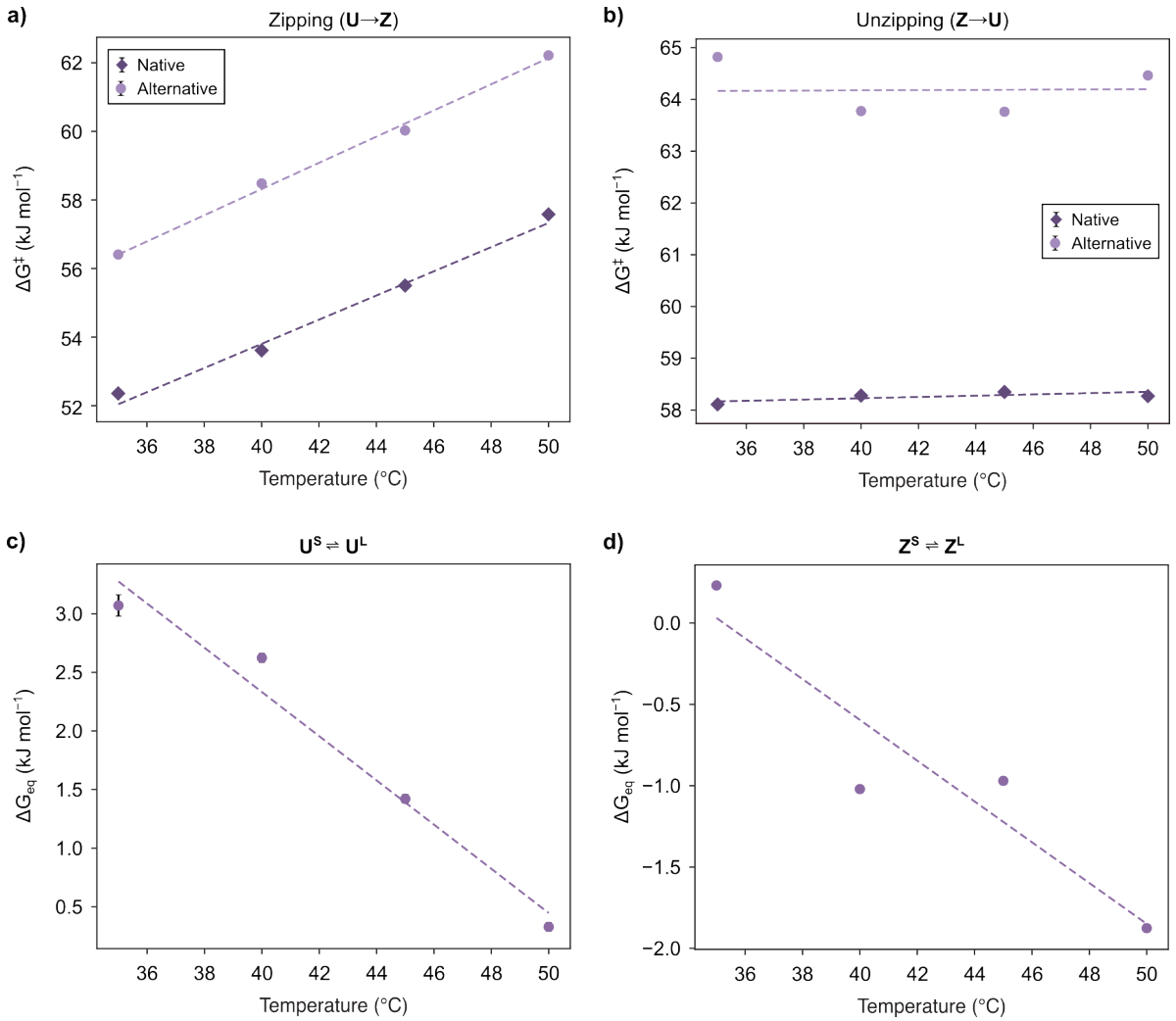

Figure S8. The activation and equilibrium free energies for the U and Z conformations. (a) Activation free energy of zipping ($\Delta G^{\ddagger}$) as a function of temperature for the native (dark purple diamonds) and alternative (light purple circles) pathways along with linear fits to the data (dashed dark purple and dashed light purple lines, respectively)**. (b)** Activation free energy of unzipping ($\Delta G^{\ddagger}$) as a function of temperature for the native (dark purple diamonds) and alternative (light purple circles) pathways along with linear fits to the data (dashed dark purple and dashed light purple lines, respectively). **(c)** Equilibrium free energy (ΔG_{eq}) for the $\boldsymbol{U^S} \rightleftharpoons \boldsymbol{U^L}$ equilibrium as a function of temperature (circles) along with a linear fit to the data (dashed line). **(d)** Equilibrium free energy (ΔG_{eq}) for the $\boldsymbol{Z^S} \rightleftharpoons \boldsymbol{Z^L}$ equilibrium as a function of temperature (circles) along with a linear fit to the data (dashed line). Error bars in all plots represent the propagated error of the fit of the double-exponential decay function to the survival probability of the dwell-time distributions.

Table S1. Sequences of the RNA stem-loop constructs.

Stem-loop Construct	Sequence
h-SL-bio	5'-hydrazide-GCUUCGGC-biotin-3'
h-SL	5'-hydrazide-GCUUCGGC-3'

Table S2. Thermodynamic parameters for the two-state model of RNA stem-loop zipping based on CD thermal melting experiments. Parameters from the fit of the set of parametric equations for two-state, unimolecular transitions between the $\boldsymbol{U}$ and $\boldsymbol{Z}$ conformations to the data from the CD thermal melting experiments. The parameters are the average of fitting results for two independently prepared samples and the errors are the standard deviations of the fitting results.

	$\Delta H_{eq}^{(2)}$ (kJ mol^{-1})	$\Delta S_{eq}^{(2)}$ (J mol^{-1}K^{-1})	$\Delta G_{eq}^{(2)}(37)$ (kJ mol^{-1})	T_m (°C)
h-SL-bio	−113(2)	−342(8)	−6.84(0.01)	57.0(0.5)
h-SL	−115.7(0.2)	−351(1)	−7.0(0.2)	56.9(0.6)

Table S3. Comparison of thermodynamic parameters determined from smFET and CD thermal melting experiments. Errors (in parentheses) for the parameters obtained from the analysis of the smFET experiments represent the standard deviations of the parameters calculated from three separate 20-s windows for the current trajectories at each temperature. Error (in parentheses) for the parameters obtained from the analysis of the CD thermal melting experiments represent the standard deviations of the parameters calculated from two independent experiments.

	$\Delta G_{eq}^{(2)}$ (kJ mol^{-1})		f_U	
	smFET	CD	smFET	CD
35°C	−8.8(0.3)	−7.520(0.003)	0.031(0.004)	0.0504(0.0007)
40°C	−5.9(0.5)	−5.81(0.04)	0.09(0.02)	0.097(0.001)
45°C	−4.1(0.2)	−4.10(0.07)	0.17(0.01)	0.175(0.004)
50°C	−2.74(0.04)	−2.4(0.1)	0.265(0.003)	0.291(0.009)

Table S4. The two kinetic phases describing the zipping of the UUCG stem-loop. Errors (in parentheses) represent the error of the fit for the double-exponential decay function to the survival probability of the dwell-time distributions of the $\boldsymbol{U}$ conformations at each temperature.

	Slow phase		**Fast phase**	
	$k_{U\to Z}^{slow}$ (s^{-1})	$A_{U\to Z}^{slow}$	$k_{U\to Z}^{fast}$ (s^{-1})	$A_{U\to Z}^{fast}$
35°C	1740(40)	0.231(0.008)	8500(100)	0.767(0.008)
40°C	1130(10)	0.267(0.004)	7360(70)	0.732(0.004)
45°C	914(8)	0.363(0.004)	5060(60)	0.622(0.004)
50°C	584(4)	0.452(0.005)	3280(50)	0.511(0.004)

Table S5. The two kinetic phases describing the unzipping of the UUCG stem-loop. Errors (in parentheses) represent the error of the fit of the double-exponential decay function to the survival probability of the dwell-time distributions of the $\boldsymbol{Z}$ conformations at each temperature.

	Slow phase		**Fast phase**	
	$k_{Z\to U}^{slow}$ (s^{-1})	$A_{Z\to U}^{slow}$	$k_{Z\to U}^{fast}$ (s^{-1})	$A_{Z\to U}^{fast}$
35°C	65.2(0.3)	0.391(0.001)	900(10)	0.428(0.003)
40°C	148.2(0.4)	0.544(0.002)	1220(10)	0.367(0.002)
45°C	222.5(0.9)	0.542(0.003)	1720(30)	0.376(0.003)
50°C	252.9(0.7)	0.629(0.002)	2540(40)	0.313(0.003)

Table S6. Activation free energies, enthalpies, and entropies of the zipping and unzipping of the UUCG stem-loop along the native and alternative pathways. Errors (in parentheses) of the $\Delta G^{\ddagger}$s are propagated from the corresponding transition rate constants. Errors for $\Delta H^{\ddagger}$s and $\Delta S^{\ddagger}$s represent the error of the fit of the linear function to the $\Delta G^{\ddagger}$s as a function of absolute temperature.

		Zipping			Unzipping		
		$\Delta \boldsymbol{G}^{\ddagger}$ (kJ mol^{-1})	$\Delta H^{\ddagger}$ (kJ mol^{-1})	$\Delta S^{\ddagger}$ (J K^{-1} mol^{-1})	$\Delta \boldsymbol{G}^{\ddagger}$ (kJ mol^{-1})	$\Delta H^{\ddagger}$ (kJ mol^{-1})	$\Delta S^{\ddagger}$ (J K^{-1} mol^{-1})
Native Pathway	**35°C**	52.36(0.04)	−57(1)	−352(3)	58.11(0.03)	54(1)	−12(3)
	40°C	53.61(0.02)			58.28(0.03)		
	45°C	55.50(0.03)			58.35(0.04)		
	50°C	57.58(0.04)			58.27(0.04)		
Alternative Pathway	**35°C**	56.41(0.06)	−61.2(0.9)	−382 (3)	64.82(0.01)	63.5(0.2)	−2.1(0.7)
	40°C	58.48(0.03)			63.775(0.007)		
	45°C	60.03(0.02)			63.76(0.01)		
	50°C	62.22(0.02)			64.463(0.007)		

Table S7. Equilibrium free energies, enthalpies, and entropies for the two kinetic phases of zipping and unzipping of the UUCG stem-loop. Errors (in parentheses) of the ΔG_{eq}s are propagated from the corresponding ratio of pre-factors for the double-exponential decay. Errors for ΔH_{eq}s and ΔS_{eq}s represent the error of the fit of the linear function to the ΔG_{eq}s as a function of absolute temperature.

		ΔG_{eq} (kJ mol^{-1})	ΔH_{eq} (kJ mol^{-1})	ΔS_{eq} (J K^{-1} mol^{-1})
$\boldsymbol{U^S \rightleftharpoons U^L}$	**35°C**	3.07(0.09)	61(8)	190(20)
	40°C	2.62(0.04)		
	45°C	1.42(0.04)		
	50°C	0.33(0.04)		
$\boldsymbol{Z^S \rightleftharpoons Z^L}$	**35°C**	0.23(0.02)	40(10)	130(30)
	40°C	−1.02(0.02)		
	45°C	−0.97(0.02)		
	50°C	−1.88(0.02)		