Salt Effects on the Thermodynamics of a Frameshifting RNA Pseudoknot under Tension

Naoto Hori,[†,‡] Natalia A. Denesyuk,[†] and D. Thirumalai[*,†,‡]

†*Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

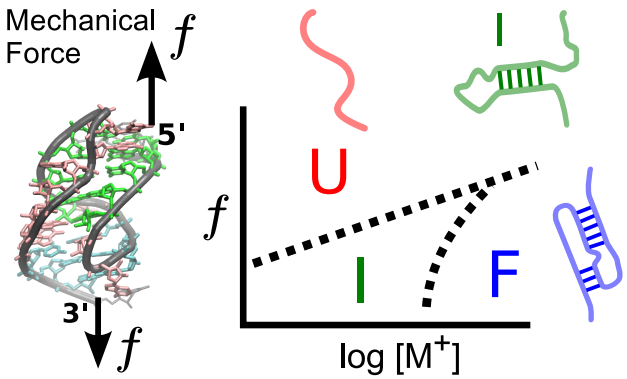
‡*Present address: Department of Chemistry, University of Texas, Austin, TX 78712, USA*

E-mail: dave.thirumalai@gmail.com

Abstract

One of the key factors in -1 programmed ribosomal frameshifting (PRF) is a pseudoknot (PK) RNA. A number of single molecule experiments have been performed on PKs to decipher the mechanism of PRF. Motivated by the experiments, we performed simulations to describe the response of a PK over a range of mechanical forces (fs) and monovalent salt concentrations (Cs). The coarse-grained simulations quantitatively reproduces the multistep thermal melting. The free energy changes obtained in simulations are in excellent agreement with experiments, thus validating our model. The predicted phase diagram shows that a sequence of structural transitions, populating distinct intermediates, occur as f and C are changed. The stem-loop tertiary interactions rupture first followed by unfolding of the 3′-end hairpin (U⇌I). Finally, the 5′-end hairpin unravels producing a stretched state (I⇌F). A theoretical analysis of the phase boundaries shows that the critical force for rupture scales as $(\log C_{\mathrm{m}})^{\alpha}$ with $\alpha = 1\,(0.5)$ for U⇌I (I⇌F) transition. This relation is used to obtain the ion-preferential coefficients, which shows that ion-RNA interactions can be quantitatively

measured in single molecule experiments. Our simulations predict a plausible link between the stability of the 5′ hairpin and efficiency of PRF.



Graphical abstract

INTRODUCTION

The multifarious roles RNA molecules play in controlling a myriad of cellular functions[1] have made it important to understand in quantitative terms their folding[2–6] and how they respond to external stresses.[7] Among them one of simplest structural motifs is the RNA pseudoknot (PK), which is involved in many biological functions. The simplest type, referred to as H-type PK, consists of two stems connected by two loops in which one of the stems forms base pairs with the loop of the other. The PK motif, found in many RNA molecules such as telomerase,[8] mRNA,[9] ribosomal RNA,[10] transfer-messenger RNA, and viral RNA,[11] are functionally important.[12] For instance, a PK in the human telomerase RNA (hTR) is essential for the enzyme (a ribonucleoprotein complex) activity.[13] Many pathogenic mutations in the hTR are found in the well-conserved PK region of the RNA further underscoring the importance of the PK.[8,13] The presence of PKs are also important in ribozyme catalysis and inhibition of ribosome translation.

Recently, there is heightened interest in the biophysics of PK folding because it plays a crucial role in affecting the efficiency of -1 programmed ribosomal frameshifting (-1 PRF).[14–16] Usually proteins are synthesized when the ribosome reads the mRNA code in three nucleotide steps until the stop codon is reached. In -1 PRF, the open reading frame of mRNA being translated within the ribosome is programmed to be shifted by one nucleotide, and consequently the mRNA becomes non-functional or produces an entirely different protein.[5,17–20] The apparent efficiency of -1 PRF depends on the presence of a seven nucleotide "slippery" sequence where frameshifting occurs. In addition, -1 PRF is also stimulated by a co-transcriptionally folded downstream structural motif, which is typically a PK. The downstream PK of the mRNA is a roadblock at the entrance to the mRNA channel on the ribosome, that impedes the translocation of the ribosome. The ribosome must unfold the PK, presumed to occur by exertion of mechanical force, to complete translocation. Because of the important role PK plays in -1 PRF and the observation that the frameshifting efficiency could depend on how the PK responds to force, a number of single molecule pulling experi-

ments have focussed on PKs.[15,21,22] A few of these studies have suggested that frameshifting efficiency is correlated with the stability of the PK based in part on the observation that structural stability modulated through nucleotide mutations affects the -1 PRF efficiency.[23,24] However, it has been also pointed out that the frameshifting efficiency cannot be explained solely based on thermodynamic or mechanical stability of PKs.[14,15,25,26] In addition, trans-acting factors have been proposed as modulators of -1 PRF efficiency,[27,28] and other, as yet unidentified, factors might also play a role. Nevertheless, given that ribosome translocation exerts mechanical force on the downstream mRNA, it is physically reasonable to assume that resistance of PK to force should at least be one factor determining the frameshifting efficiency.[14,29]

The considerations given above and the availability of both ensemble and single molecule pulling experimental data prompted us to focus on the effects of temperature, mechanical force, and salt concentration on a 28-nucleotide H-type RNA PK from the Beet Western Yellow Virus (BWYV).[19] Because of its -1 PRF property, this PK has been the subject of several ensemble experiments, which have characterized the folding thermodynamics as a function of monovalent and divalent cations.[30,31] The BWYV PK has two stems (S1 and S2) connected by two loops (L1 and L2) as shown in Figure 1. The crystal structure reveals that L1 forms triplex-type tertiary interactions with S1.[32] In another context, it has been shown that such interactions and mutations affecting them dramatically changes the efficiency of frameshifting.[21] In vitro and in vivo experiments on many mutants of BWYV suggest variations in the frameshift efficiency, which for some mutants is attributed to changes in the stem-loop interactions affecting the stability of PK. For others, direct involvement of the ribosome is invoked.[23] In addition to mechanical force (f), it is known from several ensemble experiments that the stability of BWYV PK, like other RNA molecules, depends on monovalent and divalent cations.[33] Thus, in order to link the ability of PK to stimulate -1 PRF, it is important to understand in quantitative detail how this crucial structural motif respond to f as well as changes in ion concentration.

Although single molecule pulling experiments suggest how force affects the stability of RNA,[19] structural information obtained from such experiments is limited only to changes in molecular extensions. In order to provide a theoretical basis for the changes in the stability and structural transitions as f and monovalent ion concentration (C) are altered, we performed simulations using a coarse-grained three site interaction (TIS) model[34] of the PK. After demonstrating that our simulations reproduce the temperature-induced structural transitions observed in ensemble experiments, we calculated the global phase diagrams in the $[C, f]$ plane. The simulations, performed in the same force range as used in Laser Optical Tweezer (LOT) experiments,[19] quantitatively reproduce the experimental observations. By combining our simulations and a theory based on ion-RNA interactions expressed in terms of ion-preferential coefficient,[31,35] we produce several key predictions. (1) The phase diagram in the $[C, f]$ plane shows that the folded state (F) ruptures in stages populating distinct intermediates as f increases. Structurally the low force intermediate is similar to the one populated during thermal melting. The sequence of f-induced structural transitions can be described as F$\rightleftharpoons$I$\rightleftharpoons$E where E is an extended state. (2) The critical force, f_{c}, to rupture the PK scales as $(\log C_{\mathrm{m}})^{\alpha}$ with $\alpha = 1$ for the I$\rightleftharpoons$U transition and $\alpha = 0.5$ for the I$\rightleftharpoons$F transition. This result is of general validity and is applicable to other PKs as well. The slope of the f_{c} versus $\log C_{\mathrm{m}}$ is proportional to the changes in the ion-preferential coefficient between an intermediate and unfolded states. This prediction can be used in conjunction with single molecule experiments allowing for direct measurement of a key quantity in ion-RNA interactions. (3) To the extent the downstream PK is a road block for translation, our work suggests that the ribosome should exert a force $\sim (10 - 15)$ pN[36] for unimpeded translocation along mRNA. Based in part on our work we propose that there should be a link between the stimulatory effect of frameshifting and the stability of the 5′-end of the PK. Our predictions are amenable to validations using single molecule experiments.

RESULTS AND DISCUSSION

Temperature-induced transitions in BWYV PK: In order to validate our model, we first performed temperature replica-exchange molecular dynamics (REMD) simulations at $C = 500$ mM with $f = 0$. Formations of stems (S1 and S2) and tertiary (stem-loop) interactions (L1 and L2) are characterized by hydrogen-bond energies $U_{\mathrm{HB}}^{\mathrm{S}}$ and $U_{\mathrm{HB}}^{\mathrm{L}}$, respectively. Changes in the distributions of hydrogen bond energies as the temperature is increased are shown in Figure S1 in the Supporting Information (SI). Figure S1 shows that there are two minima corresponding to two distinct states. The minimum that dominates at $T = 20°\mathrm{C}$ corresponds to the folded state in which both S1 and S2 are formed ($U_{\mathrm{HB}}^{\mathrm{S}} \approx -45$ kcal/mol). The minimum with $U_{\mathrm{HB}}^{\mathrm{S}} \approx -30$ kcal/mol at 80°C corresponds to a state in which S1 is intact (Figure S1). The free energy profile of $U_{\mathrm{HB}}^{\mathrm{L}}$ shows only one possible minimum around $U_{\mathrm{HB}}^{\mathrm{L}} \approx -20$ kcal/mol at $T = 20°\mathrm{C}$, indicating that L1 and L2 fold/unfold cooperatively as temperature changes.

We calculated the free energy profiles $G(R_\alpha) = -k_{\mathrm{B}}T \log P(R_\alpha)$ ($R_\alpha = R_{\mathrm{ee}}$, end-to-end distance, or R_{g}, radius of gyration) from the distributions of $P(R_\alpha)$, which are given in Figure S2 in the SI. The locations and the number of minima report on the thermodynamic transitions between the states (Figure 2). (1) At a low temperature, $T = 20°\mathrm{C}$, the PK is completely folded (F state) including all the structural motifs, S1, S2, L1 and L2. The existence of F state is also confirmed by the presence of a minimum at $R_{\mathrm{ee}} \approx 3$ nm (Figure 2A) and at $R_{\mathrm{g}} \approx 1.3$ nm (Figure 2B). These values are similar to those obtained directly from the crystal structure. (2) At $T = 60°\mathrm{C}$, the free energy profile for $U_{\mathrm{HB}}^{\mathrm{L}}$ shows that the tertiary interactions, L1 and L2, are no longer stably formed (Figure S1B). On the other hand, $G(R_{\mathrm{g}})$ shows two stable minima, indicating that there are two possible states (Figure 2B). In one state, both stems S1 and S2 are formed but tertiary interactions are absent, corresponding to the I1 state identified using UV absorbance and differential scanning calorimetry experiments.[30,31] The other conformation, in which only S2 is formed, has been also observed in experiments and is referred to as the I2 state. It should be noted that the

distribution of $U_{\rm HB}^{\rm S}$ also has two minima at 60°C (Figure S1A) whereas $G(R_{\rm ee})$ has only one minimum (Figure 2A). (3) At a still higher temperature $T = 80$°C, $G(R_{\rm ee})$ and $G(R_{\rm g})$ each have a minimum with S1 stably formed. This is also reflected in Figure S1, which shows a minimum at $U_{\rm HB}^{\rm S} \approx -30$ kcal/mol , and a minimum at $U_{\rm HB}^{\rm S} \approx 0$. Thus, completely unfolded conformations, U state, and the I2 state are populated. (4) At $T = 110$°C, both $U_{\rm HB}^{\rm S}$ and $U_{\rm HB}^{\rm L}$ are 0 , indicating that only the U state exists. This is also reflected in $G(R_{\rm ee})$ and $G(R_{\rm g})$, which show that the PK is completely unfolded. In both profiles (Figures 2A and 2B), the center of minimum is located at larger values than in the F state ($R_{\rm ee} \approx 5$ nm and $R_{\rm g} \approx 2$ nm).

The simulations results in Figure 2 and Figure S1 show that melting of BWYV PK undergoes through sequential transitions, F⇌I1⇌I2⇌U where I1 has the topology of the F state without significant tertiary interactions (Figure 3). The range of melting temperatures for the three transitions can be obtained from the temperature dependence of the various states $f_\alpha(T)$ (α =F, I1, I2 or U) given in Figure 2C. The I1 state has a maximum population around $T \approx 50$°C. We equate this with the melting temperature $T_{\rm m1}$ associated with the F⇌I1 transition. Similarly, $T_{\rm m2}$ for the I1⇌I2 transition is $\approx$ 70°C. The $T_{\rm m3}$ for the I2⇌U transition occurs around 90°C ($f_{\rm I2}(T_{\rm m3}) = f_{\rm U}(T_{\rm m3})$). The values of the three melting temperatures obtained in our simulations are in remarkable agreement with experiments (compare the simulation results with the values at pH = 7 in Table 1 of Nixon and Giedroc[30]).

The dependence of the stability of the F state with respect to the U state, $\Delta G_{\rm UF}(T)$ is calculated using a procedure that does not rely on any order parameter.[37] The value of $\Delta G_{\rm UF}(T = 37$°C) is -14.3 kcal/mol from simulations which is in excellent agreement with experiments. The experimental value reported by Nixon and Giedroc for $\Delta G_{\rm UF}(T = 37$°C) = -13.3 kcal/mol whereas Soto et al. estimated that $\Delta G_{\rm UF}(T = 37$°C) = -15.1 kcal/mol.[30,31] We also predict that $\Delta G_{\rm UF}(T = 25$°C) is -19.0 kcal/mol. The excellent agreement between simulations and experiments for the temperature-induced structural transitions and the $\Delta G_{\rm UF}(T)$ validates the model allowing us to predict the effect of f on BWYV PK.

Diagram of states in the $[C, f]$ plane: In order to obtain the phase diagram in the $[C, f]$ plane of the BWYV PK, we performed a series of low friction Langevin dynamics simulations by varying the salt concentration from 5 to 1200 mM and mechanical force from 0 to 20 pN, at a constant temperature of 50°C. We varied C by changing the Debye length in the simulations and the mechanical force was externally applied to the terminal nucleotides (Figure 1A).

Determination of the phase diagram in the $[C, f]$ plane requires an appropriate order parameter that can distinguish between the various states. In single molecule pulling experiments the variable that can be measured is the molecular extension, $R_{\rm ee}$, which is conjugate to the applied force. The overall size of the RNA molecule is assessed using the radius of gyration, $R_{\rm g}$. Therefore, we characterized the states of the RNA using both $R_{\rm g}$ and $R_{\rm ee}$ as order parameters. The values of the predicted $R_{\rm g}$ at $f = 0$ can be measured using scattering experiments. Using these two parameters, we obtained the $[C, f]$ phase diagram (Figure 4). Comparison of the diagram of states in Figures 4D and 4E reveal common features as well as some differences. From Figures 4D and 4E, we infer that at $f > 12.5$ pN, extended (E) conformations are dominant at all values of C. As the force decreases, the PK forms compact structures. The boundary separating the extended and relatively compact phases depends on the salt concentration as well as the value of the mechanical force. The critical force to rupture the compact structures increases linearly as a function of salt concentration (Figure 4D; boundary between red and green regions). At low forces ($f < 2.5$ pN), the diagram of states based on $R_{\rm ee}$ shows that the extension is relatively small as C changes from a low to a high value. From this finding one might be tempted to infer that the PK is always folded, which is not physical especially at low ($C \approx 10$ mM) ion concentrations. In contrast, Figure 4E shows that below 5 pN there is a transition from compact structures ($R_{\rm g} \approx 1.3$ nm in the blue region) at high C to an intermediate state ($R_{\rm g} > 2.2$ nm in the green region) at $C \approx 100$ mM. The differences between the diagrams of states in the $[C, f]$ plane using $R_{\rm ee}$ and $R_{\rm g}$ as order parameters are more vividly illustrated in terms of the free

energy profiles $G(R_\alpha) = -k_\mathrm{B}T\log(R_\alpha)$ where R_α is R_ee or R_g (Figure 5) The profiles $G(R_\mathrm{ee})$, at three values of C and different f values, show at most two minima. At $f = 0$, there is only one discernible minimum at $R_\mathrm{ee} \approx 4.2$ nm at $C = 10$ mM. The minimum moves to $R_\mathrm{ee} \approx 3$ nm at $C = 1200$ mM corresponding to a folded PK. At $f = 5$ pN there are two minima at $C = 10$ mM corresponding to a compact structure and a stretched conformation (see the cyan profile in Figure 5A). As f exceeds 5 pN, there is essentially one minimum whose position increases as f increases. In the force regime ($f > 5$ pN) only the E state is visible at all C.

A much richer diagram of states emerges when the $G(R_\mathrm{g}) = -k_\mathrm{B}T\log P(R_\mathrm{g})$ is examined. The $G(R_\mathrm{g})$ profile at $C = 10$ mM is similar to $G(R_\mathrm{ee})$. However, at higher concentrations there is clear signature of three states (especially at $f = 12.5$ pN) corresponding to the folded PK, an intermediate state, and the E state. The free energy profiles show that under force there are minimally three states, which are F, I, and E. Because R_ee cannot capture the complexity of the states populated as C and f are varied, it is important to consider R_g as well to fully capture the phase diagram of the PK.

Formation of stems and loops: In order to highlight the structural changes that occur as C and f are varied, we used, Q, the fraction of native hydrogen-bond interactions as an order parameter. The $[C, f]$ phase diagram (Figure 6A) calculated using the average Q is similar to the R_g-phase diagram (Figure 4E), indicating the presence of three states (Figure 6A). Using Q we can also quantitatively assess the contributions from different structural elements to the stability of the PK, and correctly infer the order in which the structural elements of the PK rupture as f is increased. In order to determine the structural details of each state, we calculated the individual Qs for the two stems (Q_S1, Q_S2) and the two loops (Q_L1 Q_L2) (Figure 6B). The dependence of $\langle Q_\mathrm{S1}\rangle$ as a function of C and f shows that Stem 1 (S1) is extremely stable and remains intact at all salt concentration, rupturing only at $f \approx 10$ pN (see upper left panel in Figure 6B). In contrast, the upper right panel in

Figure 6B shows that Stem 2 (S2) is only stably folded above a moderate salt concentration ($C \approx 80$ mM) and low f. The stability difference between S1 and S2 can be explained by the number of canonical G-C base pairs; S1 has five G-C base pairs whereas only three such pairs are in S2. Consequently, S1 cannot be destabilized even at low C, but its rupture to produce extended states requires the use of mechanical force. Above $C > 100$ mM, the fraction of contacts associated with S2 and the two loops (Q_{S2}, Q_{L1} and Q_{L2}) simultaneously increase. All the interactions are, however, still not completely formed ($Q \approx 0.8$) even at the highest salt concentration, $C = 1200$ mM. In particular, the contacts involving L2, $Q_{\mathrm{L2}} \approx 0.6$, implying that tertiary interactions are only marginally stable at $T = 50°$C.

The results in Figure 6 provide a structural explanation of the transitions that take place as C and f are changed. (1) In the high force regime, the two stems are fully ruptured and native interactions are absent creating the E state. (2) Below $C \approx 100$ mM and $f \lesssim 10$ pN (left bottom on the $[C, f]$ plane), only S1 is intact, and thus a hairpin conformation is dominant. This is the intermediate state, I2, identified by Soto et al.[31] (3) Above $C \gtrsim 100$ mM and $f \lesssim 10$ pN (right bottom), both S1 and S2 stems are formed and the tertiary interactions involving the two loops are at least partly formed, ensuring that the topology of the PK is native-like. At $f = 0$ and temperature as a perturbation, it has been shown that another intermediate state, I1, exist between I2 and folded states.[30,31] The I1 state may correspond to two stems fully folded with the entire topology of the native conformation intact, but one in which tertiary interactions between stems and loops are incomplete. Although we have accounted for the presence of such a state in our simulations (see Figure 2) from here on we do not distinguish between I1 and the completely folded state in the phase diagram since the increases in Q_{S2}, Q_{L1} and Q_{L2} on the $[C, f]$ plane almost overlap. Therefore, we classify the right-bottom region in the $[C, f]$ phase diagram as (I1+F) state.

Population analysis of the secondary and tertiary interactions: The $[C, f]$ phase diagrams provide a thermodynamic basis for predicting the structural changes that occur

as f and C are varied. However, they capture only the average property of the ensemble within each state without the benefit of providing a molecular picture of the distinct states, which can only be described using simulations done at force values close to those used in laser optical tweezer (LOT) experiments. Therefore, we investigated the distribution of interactions to ascertain the factors that stabilize the three states as C and f are varied. Our analysis is based on the hydrogen bond (HB) energy, U_{HB}. The energy function for HB captures the formation of base pairs and loop-stem tertiary interactions, taking into account both distance and angular positions of the donor and the accepter.[37]

Figure 7 shows that the probability densities of the total HB energy have three peaks at $U_{\mathrm{HB}} \approx 0$, -30, and -60 kcal/mol. In the high-force regime ($f > 12.5$ pN), there is only one population at all C around $U_{\mathrm{HB}} \approx 0$, which obviously corresponds to the fully extended state. The total HB energy can be decomposed into contributions arising from four distinct structural motifs in the PK (Figures S3A-D). The peak at $U_{\mathrm{HB}} \approx -30$ kcal/mol (Figure 7) arises due to the formation of S1 (compare Figure 7 and S3A). The broad peak at $U_{\mathrm{HB}} \approx -60$ kcal/mol (Figure 7A) is due to the sum of contributions from S2 and the tertiary interactions.

Parsing the U_{HB} due to contributions from interactions among S1, S2, L1, and L2 produces a picture of thermodynamic ordering as C (f) is increased (decreased) from low (high) value to high (low) value. Low C and high f correspond to the top left region in the phase diagrams (Figures 4 and 6). In all conditions, S1 is the first to form as indicated by the presence of distinct peaks at $C > 100$ mM and $f \lesssim 10$ pN in Figure S3A. Under more favorable conditions ($f < 10$ pN; $C \approx 300$ mM or greater) S2 and tertiary interactions involving L1 are established (Figures S3 B and C). Only at the highest value of C and $f \lesssim 5$ pN, tertiary interactions involving L2 emerge. Interestingly, the order of formation of the various structural elements en route to the formation of their folded state follows the stabilities of the individual elements with the most stable motif forming first. The PK tertiary interactions are consolidated only after assembly of the stable structural elements. This finding follows from our earlier work establishing that the hierarchical assembly of RNA PKs is determined

by the stabilities of individual structural elements.[38]

We did not find any condition in which S1 and S2 are completely folded but there are no tertiary interactions, $U_{\rm HB}^{\rm L} \approx 0$ (i.e. isolated I1 state). Our simulations show that the tertiary interactions contribute to the loop energy only after the stems are formed. Thus, the actual population is distributed between I1 and F, and the tertiary interactions are only marginally stable in the conditions examined here. Soto et al. suggested that diffuse Mg^{2+} ions play an important role in stabilizing such tertiary interactions.[31] Because Mg^{2+} is not considered here, we cannot determine the precise interactions stabilizing the I1 state, especially at $f \neq 0$. In the discussion hereafter, for simplicity, we refer to (I1+F) state as F.

Critical rupture force and ion preferential interaction coefficient: In order to delineate the phase boundaries quantitatively, we calculated the free energy differences between the three major states in the $[C, f]$ plane using $\Delta G_{\rm UI}(C, f) = -k_{\rm B}T \log \frac{P({\rm I};C,f)}{P({\rm U};C,f)}$, where the classification of the states, U, I (=I2), or F, is based on the hydrogen-bond interaction energies of stems, $U_{\rm HB}^{\rm S1}$ and $U_{\rm HB}^{\rm S2}$ (Figure 8). A similar equation is used to calculate $\Delta G_{\rm IF}$. Based on the distribution of $U_{\rm HB}$, we determined the threshold values for S1 formation as $U_{\rm HB}^{\rm S1} < -15\,$kcal/mol (Figure S3A). Similarly, S2 formation is deemed to occur if $U_{\rm HB}^{\rm S2} < -10\,$kcal/mol (Figure S3B). We classified each structure depending on whether both stems are formed (F) with both $U_{\rm HB}^{\rm S1} < -15$ kcal/mol and $U_{\rm HB}^{\rm S2} < -10$ kcal/mol, only S1 is formed (I(=I2)), or unfolded (U).

The classification of the diagram of states based on the calculation of $\Delta G_{\rm UI}$ and $\Delta G_{\rm IF}$ is consistent with the $R_{\rm g}$ and Q phase diagrams (Figures 8C and 8D). At all values of f, ΔG depends linearly on $\log C$ over a broad range of C as shown in Figures 8A and 8B. The midpoints of salt concentration, $C_{\rm m}$s, at each f, determined using $\Delta G_{\rm UI}(C_{\rm m}^{\rm UI}, f) = 0$ and $\Delta G_{\rm IF}(C_{\rm m}^{\rm IF}, f) = 0$ leads to $f_{\rm c}^{\rm UI}(C_{\rm m}^{\rm UI})$ and $f_{\rm c}^{\rm IF}(C_{\rm m}^{\rm IF})$, respectively, for the U$\rightleftharpoons$I and I$\rightleftharpoons$F transitions. The forces $f_{\rm c}^{\rm UI}(C_{\rm m}^{\rm UI})$ and $f_{\rm c}^{\rm IF}(C_{\rm m}^{\rm IF})$ are the critical forces needed to rupture the I and F states, respectively. The locus of points $f_{\rm c}^{\rm UI}(C_{\rm m})$ and $f_{\rm c}^{\rm IF}(C_{\rm m})$ are shown as circles

in Figures 8C and 8D, respectively. It is clear that $f_{\rm c}^{\rm UI}(C_{\rm m}^{\rm UI})$ is linear in $\log C_{\rm m}^{\rm UI}$ along the phase boundary. In contrast, the dependence of $f_{\rm c}^{\rm IF}(C_{\rm m}^{\rm IF})$ on $\log C_{\rm m}^{\rm IF}$ is non-linear. In what follows we provide a theoretical interpretation of these results, which automatically shows that the difference in ion preferential coefficients between states can be directly measured in LOT experiments.

The salt-concentration dependence of RNA stability can be determined using the preferential interaction coefficient, defined as $\Gamma = \left(\frac{\partial C}{\partial C_{\rm RNA}}\right)_{\mu}$, where $C_{\rm RNA}$ is the concentration of RNA, and the chemical potential of the ions, μ, is a constant.[35,39] The free energy difference, $\Delta G_{\alpha\beta}$ ($=G_\beta - G_\alpha$), between two states α and β (such as U and I or I and F) can be expressed as,

$$\Delta G_{\alpha\beta}\left(C\right) = \Delta G_{\alpha\beta}\left(C_0\right) - 2k_{\rm B}T\Delta\Gamma_{\alpha\beta}\log\frac{C}{C_0}, \tag{1}$$

where C_0 is an arbitrary reference value of the salt concentration.[40] Note that we consider only 1:1 monovalent salt such as KCl or NaCl for which ion activity can be related to salt concentration C. The factor of 2 in Eq. (1) arises from charge neutrality. The difference, $\Delta\Gamma_{\alpha\beta}$ ($=\Gamma_\beta - \Gamma_\alpha$), is interpreted as an effective number of bound cations or excluded anions when the RNA molecule changes its conformation between the two states.[35,39] The free energy change upon application of an external force f is $\Delta G_{\alpha\beta}(f) = \Delta G_{\alpha\beta}(0) + \Delta R_{\rm ee}^{\alpha\beta}\cdot f$ where $\Delta R_{\rm ee}^{\alpha\beta} = R_{\rm ee}^{\alpha} - R_{\rm ee}^{\beta}$.[29] Thus,

$$\begin{aligned}\Delta G_{\alpha\beta}(C,f) &= \Delta G_{\alpha\beta}(C_0,0) \\ &\quad -2k_{\rm B}T\Delta\Gamma_{\alpha\beta}\log\frac{C}{C_0} + \Delta R_{\rm ee}^{\alpha\beta}\cdot f.\end{aligned} \tag{2}$$

In the $[C,f]$ phase diagram, $\Delta G_{\alpha\beta} = 0$ along the phase boundaries. Since the reference salt concentration C_0 is arbitrary, we determine its value using $\Delta G_{\alpha\beta}(C_0, f=0) = 0$. Thus, $C_0 = C_{{\rm m},f=0}^{\alpha\beta}$ is the salt concentration at the midpoint of the transition at zero force. With

this choice of C_0, Eq.(2) can be adopted to obtain the phase boundary at $f \neq 0$ using,

$$0 = -2k_{\mathrm{B}}T\Delta\Gamma_{\alpha\beta} \log \frac{C_{\mathrm{m}}^{\alpha\beta}}{C_{\mathrm{m},f=0}^{\alpha\beta}} + \Delta R_{\mathrm{ee}}^{\alpha\beta} \cdot f_{\mathrm{c}}^{\alpha\beta}, \tag{3}$$

where $C_{\mathrm{m}}^{\alpha\beta}$ is the midpoint of salt concentration, and $f_{\mathrm{c}}^{\alpha\beta}$ is the critical force associated with the transition between states α and β. A relation that is closely related to Eq. (3) was derived recently by Saleh and coworkers[41] who also discussed the applicability of their approach based on Clausius-Clapeyron equation away from the phase boundary ($\Delta G_{\alpha\beta} = 0$). It follows from Eq.(3) that the critical force,

$$f_{\mathrm{c}}^{\alpha\beta} = \frac{2k_{\mathrm{B}}T\Delta\Gamma_{\alpha\beta}}{\Delta R_{\mathrm{ee}}^{\alpha\beta}} \log \frac{C_{\mathrm{m}}^{\alpha\beta}}{C_{\mathrm{m},f=0}^{\alpha\beta}}, \tag{4}$$

leading to the prediction that there should be a linear-dependence between $f_{\mathrm{c}}^{\alpha\beta}$ and the logarithm of the midpoint of the salt concentration if $\Delta R_{\mathrm{ee}}^{\alpha\beta}$ is a constant value along the phase boundary. In accord with the prediction in Eq.(4), we find that $f_{\mathrm{c}}^{\mathrm{UI}}$ varies linearly with $\log \frac{C_{\mathrm{m}}^{\mathrm{UI}}}{C_{\mathrm{m},f=0}^{\mathrm{UI}}}$ along the phase boundary separating U and I (Figure 8C) where $\langle \Delta R_{\mathrm{ee}}^{\mathrm{UI}} \rangle \approx 4.9\,\mathrm{nm}$ over a broad range of f_{c} (Figures 9A and 9C).

The observed non-linearity of the phase boundary separating I and F over the entire concentration range (Figure 8D) arises because $\Delta R_{\mathrm{ee}}^{\mathrm{IF}}$ is not a constant, but varies linearly with $f_{\mathrm{c}}^{\mathrm{IF}}$ (Figures 9B and 9C). Thus, using $\Delta R_{\mathrm{ee}}^{\alpha\beta} = a f_{\mathrm{c}}^{\alpha\beta} + b$ and Eq.(4), we find that $f_{\mathrm{c}}^{\alpha\beta}$ satisfies,

$$f_{\mathrm{c}}^{\alpha\beta} = \frac{2k_{\mathrm{B}}T\Delta\Gamma_{\alpha\beta}}{a f_{\mathrm{c}}^{\alpha\beta} + b} \log \frac{C_{\mathrm{m}}^{\alpha\beta}}{C_{\mathrm{m}.f=0}^{\alpha\beta}}. \tag{5}$$

Note that if a is zero then Eq.(5) reduces to Eq.(4) with $b = \Delta R_{\mathrm{ee}}^{\alpha\beta} = const$. Solving Eq.(5), the nonlinear dependence of $f_{\mathrm{c}}^{\mathrm{IF}}$ is expressed as,

$$f_{\mathrm{c}}^{\mathrm{IF}} = \sqrt{\frac{2k_{\mathrm{B}}T\Delta\Gamma_{\mathrm{IF}}}{a} \log \frac{C_{\mathrm{m}}^{\mathrm{IF}}}{C_{\mathrm{m},f=0}^{\mathrm{IF}}} + \frac{b^2}{4a^2}} - \frac{b}{2a}. \tag{6}$$

The simulation data can be quantitatively fit using Eq.(6) (Figure 8D). In general, we expect

$\Delta R_{\rm ee}^{\alpha\beta}$ depends on $f_{\rm c}^{\alpha\beta}$, and hence $f_{\rm c}^{\alpha\beta} \sim \sqrt{\log C_{\rm m}^{\alpha\beta}}$. It is worth noting that estimation of $\Delta\Gamma_{\alpha\beta}$ requires only knowledge of how the critical force varies with salt concentrations ($\Delta G = 0$). For any given salt concentration the best statistics are obtained at the critical force because at $f_{\rm c}$ multiple transitions between the two states can be observed.

From the coefficients in the dependence of $f_{\rm c}$ on $\log C_{\rm m}$, we can estimate $\Delta\Gamma$ which provides a quantitative estimation of the extent of ion-RNA interactions.[40] For the U⇌I transition, $\langle \Delta R_{\rm ee} \rangle \approx 4.9\,$nm and the slope of the linear function is 1.7 pN which leads to $\Delta\Gamma_{\rm UI} = 0.96$ (Figure 8C). This indicates that, upon the conformational change from unfolded to the intermediate state, one cation (anion) is effectively bound (excluded) to the PK. For the transition between I and F, fitting the dependence of $\Delta R_{\rm ee}^{\rm IF}$ on $f_{\rm c}^{\rm IF}$ with $a = 0.14$ nm/pN and $b = 0.27$ nm (Figure 9C), we estimate $\Delta\Gamma_{\rm IF} = 1.8$ using Eq.(6) (Figure 8D). Consequently, the total number of excess cations effectively bound to folded RNA is expected to be around 2.8, a value that is in rough accord with experimental estimate of 2.2.[30] More importantly, $\Delta\Gamma_{\alpha\beta}$ can be measured using LOT experiments as shown recently, thus expanding the scope of such experiments to measure ion-RNA interactions.

MODEL AND METHODS

The three-interaction-site (TIS) model: We employed a variant of the three-interaction-site (TIS) model, a coarse-grained model first introduced by Hyeon and Thirumalai for simulating nucleic acids.[34] The TIS model has been previously used to make several quantitative predictions for RNA molecules ranging from hairpins to ribozymes with particular focus on folding and response to f.[34,38,42–44] More recently, we have incorporated the consequences of counter ion condensation into the TIS model, allowing us to predict thermodynamic properties of RNA hairpins and PK that are in remarkable agreement with experiments.[37] In the TIS model, each nucleotide is represented by three coarse-grained spherical beads corresponding to phosphate (P), ribose sugar (S), and a base (B). Briefly, the effective po-

tential energy (for details see Denesyuk and Thirumalai[37]) of a given RNA conformation is $U_{\mathrm{TIS}} = U_{\mathrm{L}} + U_{\mathrm{EV}} + U_{\mathrm{ST}} + U_{\mathrm{HB}} + U_{\mathrm{EL}}$, where U_{L} accounts for chain connectivity and angular rotation of the polynucleic acids, U_{EV} accounts for excluded volume interactions of each chemical group, U_{ST} and U_{HB} are the base-stacking and hydrogen-bond interactions, respectively. Finally, the term U_{EL} corresponds to electrostatic interactions between phosphate groups.

The repulsive electrostatic interactions between the phosphate groups is taken into account through the Debye-Hückel theory, $U_{\mathrm{EL}} = \sum_{i,j} \frac{q^{*2}e^2}{4\pi\varepsilon_0\varepsilon(T)r_{ij}} \exp\left(-\frac{r_{ij}}{\lambda_{\mathrm{D}}}\right)$, where the Debye length is $\lambda_{\mathrm{D}} = \sqrt{\frac{\varepsilon(T)k_{\mathrm{B}}T}{4\pi e^2 I}}$. In the simulations, salt concentration can be varied by changing the ionic strength $I = \Sigma q_n^2 \rho_n$ where q_n is the charge of ion of type n and ρ_n is its number density. Following our earlier study,[37] we use an experimentally fitted function for the temperature-dependent dielectric constant $\varepsilon(T)$.[45] The renormalized charge on the phosphate group is $-q^*e\,(q^* < 1)$. Because of the highly charged nature of the polyanion, counterions condense onto the RNA, thus effectively reducing the effective charge per phosphate. The extent of charge reduction can be calculated using the Oosawa-Manning theory. Charge renormalization, first used in the context of RNA folding by Heilman-Miller et al.[46] and more recently incorporated into CG models for RNA by us[37] and others,[47,48] is needed to predict accurately the thermodynamics and kinetics of RNA folding.[49] The renormalized value of the charge on the P group is approximately given by $-q^*(T)e = \frac{-be}{l_{\mathrm{B}}(T)}$, where the Bjerrum length is $l_{\mathrm{B}}(T) = \frac{e^2}{\varepsilon(T)k_{\mathrm{B}}T}$ and b is the mean distance between the charges on the phosphate groups.[50] The mean distance (b) between charges on RNA is difficult to estimate (except for rod-like polyelectrolytes) because of its shape fluctuations. Therefore, it should be treated as an adjustable parameter. In our previous study we showed that $b = 4.4$ Å provides an excellent fit of the thermodynamics of RNA hairpins and the MMTV pseudoknot.[37] We use the same value in the present study of BWYV as well. Thus, the force-field used in this study is the same as in our previous study attesting to its transferability for simulating small RNA molecules.

The native structure of the BWYV PK is taken from the PDB (437D).[32] The structure contains 28 nucleotides. We added an additional guanosine monophosphate to both the 5′ and 3′ terminus. Thus, the simulated PK has 30 nucleotides. The mechanical force is applied to the ribose sugars of both the 5′- and 3′- terminus (Figure 1A).

Simulation details: We performed Langevin dynamics simulations by solving the equation of motion, $m\ddot{\boldsymbol{x}} = -\frac{\partial U_{\mathrm{TIS}}}{\partial \boldsymbol{x}} - \gamma\dot{\boldsymbol{x}} + \boldsymbol{R}$, where m is mass, γ, the friction coefficient, depends on the size of the particle type (P, S, or B), and $\boldsymbol{R}$ is a Gaussian random force which satisfies the fluctuation-dissipation relation. The $[C, f]$ phase diagram is calculated at a somewhat elevated temperature, $T = 50°\mathrm{C}$, in order to observe multiple folding-unfolding transitions when the monovalent salt concentration is varied over a wide range, from 5 to 1200 mM. The numerical integration is performed using the leap-frog algorithm with time step length $\delta t_L = 0.05\tau$ where τ is the unit of time. We performed low friction Langevin dynamics simulations to enhance the rate of conformational sampling.[51] For each condition (salt concentration (C) and mechanical force (f)), we performed 50 independent simulations for 6×10^8 time steps and the last two third of trajectories are used for data analyses. In addition, we also performed a set of simulations using the replica-exchange molecular dynamics (REMD) to confirm the validity of the results obtained by the low friction Langevin dynamics simulations.[52] In the REMD simulations, both C and f are treated as replica variables and are exchanged between 256 replicas, by covering C from 1 to 1200 mM and f from 0 to 20 pN. The replicas are distributed over 16×16 grid. The combined set of simulations using different protocols assures us that the simulations are converged at all conditions.

Analysis of simulation data: To analyze the structural transitions from which the phase diagrams are determined we stored snapshots every 10^4 time steps from all the trajectories. We computed the radius of gyration (R_{g}), and the end-to-end distance (R_{ee}) using the stored snapshots. The fraction of native contacts, Q, is calculated by counting the number of hydrogen-bond (HB) interaction pairs. The assessment of HB formation is based on the

instantaneous value of the HB interaction energy, $U_{\rm HB}$. In our model, each HB interaction pair contributes up to $nU^0_{\rm HB}$ towards stability, where $U^0_{\rm HB}$ is -2.43 kcal/mol corresponding to the stability of one hydrogen bond and n represents number of hydrogen bond associated in the interaction.[37] A cut-off value $nU^{\rm C}_{\rm HB}$ is defined to pick out contacts that are established. We use a value, $U^{\rm C}_{\rm HB} = -1.15$ kcal/mol to obtain the diagram using Q as an order parameter (see below). Modest changes in $U^{\rm C}_{\rm HB}$ do not affect the results qualitatively.

CONCLUDING REMARKS

Motivated by the relevance of force-induced transitions in PK to programmed ribosomal frameshifting, we have conducted extensive simulations of BWYV PK using a model of RNA, which predicts with near quantitative accuracy the thermodynamics at $f = 0$. The phase diagram in the $[C, f]$ plane, which can be measured using single molecule LOT pulling experiments, shows that the thermodynamics of rupture as f is increased or the assembly of the PK as f is decreased, involve transitions between the extended, intermediate, and the folded states. The predicted linear relationship between $f_{\rm c}$ and $\log C_{\rm m}$ (Eq.4) or $f_{\rm c} \sim \sqrt{\log C_{\rm m}}$ (Eq.6) show that $\Delta\Gamma$ can be directly measured in LOT experiments. The theoretical arguments leading to Eqs.(4) and (6) are general. Thus, f can be used as a perturbant to probe ion-RNA interactions through direct measurement of $\Delta\Gamma$.

ACKNOWLEDGEMENTS

This work was completed when the authors were at the University of Maryland. We are grateful to Jon Dinman, Xin Li, Pavel Zhuravlev, Huong Vu, and Mauro Mugnai for comments on the manuscript. This work was supported in part by a grant from the National Science Foundation (CHE-1361946).

References

(1) Cech, T. R.; Steitz, J. A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **2014**, *157*, 77–94.

(2) Tinoco, I.; Bustamante, C. How RNA folds. *J. Mol. Biol.* **1999**, *293*, 271–281.

(3) Woodson, S. A. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **2005**, *15*, 324–330.

(4) Thirumalai, D.; Hyeon, C. RNA and protein folding: common themes and variations. *Biochemistry* **2005**, *44*, 4957–70.

(5) Chen, S.-J. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **2008**, *37*, 197.

(6) Woodson, S. A. RNA folding pathways and the self-assembly of ribosomes. *Acc. Chem. Res.* **2011**, *44*, 1312–1319.

(7) Tinoco, I.; TX Li, P.; Bustamante, C. Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.* **2006**, *39*, 325–360.

(8) Theimer, C. A.; Feigon, J. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **2006**, *16*, 307–18.

(9) Brierley, I.; Digard, P.; Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **1989**, *57*, 537–547.

(10) Powers, T.; Noller, H. F. A functional pseudoknot in 16S ribosomal RNA. *EMBO J.* **1991**, *10*, 2203.

(11) Ten Dam, E.; Pleij, K.; Draper, D. Structural and functional aspects of RNA pseudoknots. *Biochemistry* **1992**, *31*, 11665–11676.

(12) Staple, D. W.; Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005**, *3*, e213.

(13) Gilley, D.; Blackburn, E. H. The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6621–6625.

(14) Green, L.; Kim, C.-H.; Bustamante, C.; Tinoco, I. Characterization of the mechanical unfolding of RNA pseudoknots. *J. Mol. Biol.* **2008**, *375*, 511–528.

(15) Ritchie, D. B.; Foster, D. A. N.; Woodside, M. T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16167–16172.

(16) Kim, H.-K.; Liu, F.; Fei, J.; Bustamante, C.; Gonzalez, R. L.; Tinoco, I. A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5538–5543.

(17) Chen, G.; Wen, J.-D.; Tinoco, I. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA* **2007**, *13*, 2175–2188.

(18) Giedroc, D. P.; Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* **2009**, *139*, 193–208.

(19) White, K. H.; Orzechowski, M.; Fourmy, D.; Visscher, K. Mechanical unfolding of the beet western yellow virus -1 frameshift signal. *J. Am. Chem. Soc.* **2011**, *133*, 9775–9782.

(20) Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *WIREs RNA* **2012**, *3*, 661–673.

(21) Chen, G.; Chang, K.-Y.; Chou, M.-Y.; Bustamante, C.; Tinoco, I. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12706–11.

(22) de Messieres, M.; Chang, J.-C.; Belew, A. T.; Meskauskas, A.; Dinman, J. D.; La Porta, A. Single-molecule measurements of the CCR5 mRNA unfolding pathways. *Biophys. J.* **2014**, *106*, 244–252.

(23) Kim, Y.-G.; Su, L.; Maas, S.; O'Neill, A.; Rich, A. Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14234–14239.

(24) Mouzakis, K. D.; Lang, A. L.; Vander Meulen, K. A.; Easterday, P. D.; Butcher, S. E. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res.* **2013**, *41*, 1901–1913.

(25) Napthine, S.; Liphardt, J.; Bloys, A.; Routledge, S.; Brierley, I. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J. Mol. Biol.* **1999**, *288*, 305–320.

(26) Kontos, H.; Napthine, S.; Brierley, I. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell Biol.* **2001**, *21*, 8657–8670.

(27) Belew, A. T.; Meskauskas, A.; Musalgaonkar, S.; Advani, V. M.; Sulima, S. O.; Kasprzak, W. K.; Shapiro, B. A.; Dinman, J. D. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* **2014**, *512*, 265–269.

(28) Li, Y.; Treffers, E. E.; Napthine, S.; Tas, A.; Zhu, L.; Sun, Z.; Bell, S.; Mark, B. L.; van Veelen, P. A.; van Hemert, M. J.; Firthe, A. E.; Brierleye, I.; Snijderc, E. J.; Fang, Y. Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E2172–E2181.

(29) Tionoco, I.; Bustamante, C. The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys. Chem.* **2002**, *101*, 513–533.

(30) Nixon, P. L.; Giedroc, D. P. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.* **2000**, *296*, 659–71.

(31) Soto, A. M.; Misra, V.; Draper, D. E. Tertiary structure of an RNA pseudoknot is stabilized by "diffuse" Mg^{2+} ions. *Biochemistry* **2007**, *46*, 2973–83.

(32) Su, L.; Chen, L.; Egli, M.; Berger, J. M.; Rich, A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Mol. Biol.* **1999**, *6*, 285–292.

(33) Tan, Z.-J.; Chen, S.-J. Salt contribution to RNA tertiary structure folding stability. *Biophys. J.* **2011**, *101*, 176–87.

(34) Hyeon, C.; Thirumalai, D. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6789–94.

(35) Anderson, C. F.; Record Jr, M. T. Salt dependence of oligoion-polyion binding: a thermodynamic description based on preferential interaction coefficients. *J. Phys. Chem.* **1993**, *97*, 7116–7126.

(36) Liu, T.; Kaplan, A.; Alexander, L.; Yan, S.; Wen, J.-D.; Lancaster, L.; Wickersham, C. E.; Fredrick, K.; Noller, H.; Tinoco, I.; Bustamante, C. J. Direct measurement of the mechanical work during translocation by the ribosome. *eLife* **2014**, *3*, e03406.

(37) Denesyuk, N. A.; Thirumalai, D. Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901–11.

(38) Cho, S. S.; Pincus, D. L.; Thirumalai, D. Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17349–54.

(39) Bond, J. P.; Anderson, C. F.; Record Jr, M. T. Conformational transitions of duplex

and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients. *Biophys. J.* **1994**, *67*, 825.

(40) Record Jr, M. T.; Zhang, W.; Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem* **1998**, *51*, 281–353.

(41) Dittmore, A.; Landy, J.; Molzon, A. A.; Saleh, O. A. Single-Molecule Methods for Ligand Counting: Linking Ion Uptake to DNA Hairpin Folding. *J. Am. Chem. Soc.* **2014**, *136*, 5974–5980.

(42) Biyun, S.; Cho, S. S.; Thirumalai, D. Folding of human telomerase RNA pseudoknot using ion-jump and temperature-quench simulations. *J. Am. Chem. Soc.* **2011**, *133*, 20634–43.

(43) Lin, J.-C.; Thirumalai, D. Kinetics of allosteric transitions in S-adenosylmethionine riboswitch are accurately predicted from the folding landscape. *J. Am. Chem. Soc.* **2013**, *135*, 16641–50.

(44) Denesyuk, N. A.; Thirumalai, D. How do metal ions direct ribozyme folding? *Nature Chem.* **2015**,

(45) Malmberg, C. G.; Maryott, A. A. Dielectric Constant of Water from 0° to 100° C. *J. Res. Nat. Bur. Stand.* **1956**, *56*.

(46) Heilman-Miller, S. L.; Thirumalai, D.; Woodson, S. A. Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations. *J. Mol. Biol.* **2001**, *306*, 1157–1166.

(47) Hayes, R. L.; Noel, J. K.; Whitford, P. C.; Mohanty, U.; Sanbonmatsu, K. Y.;

Onuchic, J. N. Reduced model captures Mg^{2+}–RNA interaction free energy of riboswitches. *Biophys. J.* **2014**, *106*, 1508–1519.

(48) Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. Generalized Manning condensation model captures the RNA ion atmosphere. *Phys. Rev. Lett.* **2015**, *114*, 258105.

(49) Heilman-Miller, S. L.; Pan, J.; Thirumalai, D.; Woodson, S. A. Role of counterion condensation in folding of the Tetrahymena ribozyme II. Counterion-dependence of folding kinetics. *J. Mol. Biol.* **2001**, *309*, 57–68.

(50) Manning, G. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions I. Colligative Properties. *J. Chem. Phys.* **1969**, *51*, 924.

(51) Honeycutt, J. D.; Thirumalai, D. The nature of folded states of globular proteins. *Biopolymers* **1992**, *32*, 695–709.

(52) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

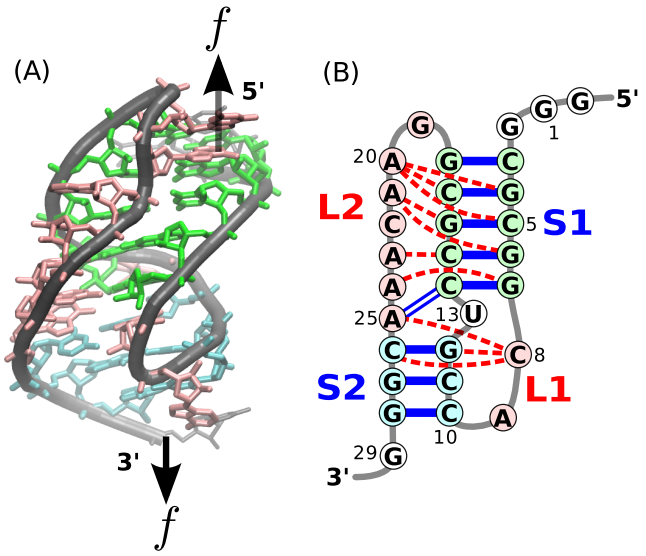


Figure 1: Structure of the BWYV pseudoknot. **(A)** Sketch of the crystal structure[32] color-coded according to secondary structures. The black arrows indicate positions at which the external mechanical force is applied. **(B)** Schematic representation of the secondary structure. Unlike the typical H-type pseudoknot, the two helix stems (S1 and S2) are not coaxially stacked in the BWYV pseudoknot.[23] Tertiary interactions are indicated by red-dashed lines.

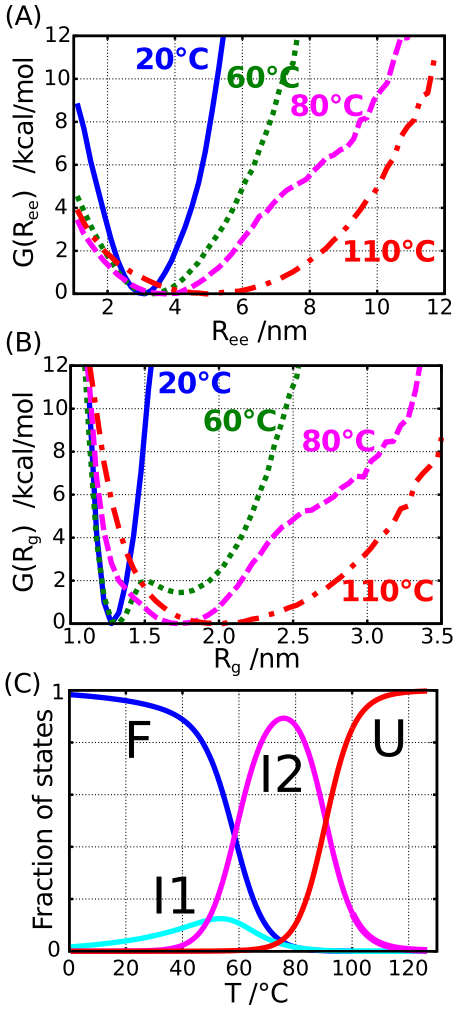


Figure 2: Characteristics of the thermally-induced folding-unfolding transition of BWYV PK simulated using the temperature-REMD method at $C = 500$ mM and $f = 0$. (A, B) Free energy profiles of the molecular extension (A; R_{ee}) and radius of gyration (B; R_{g}) at 20, 60, 80, and 110°C. Each curve is normalized to $G = 0$ at its probability $P(R)$ maximum. (C) Temperature dependence of the fraction of the four prominent states, which are explicitly labeled.

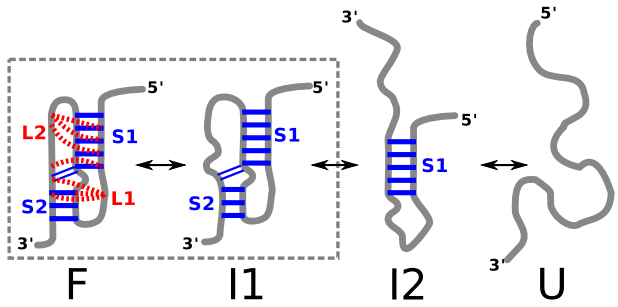


Figure 3: Schematic of the thermodynamic assembly of BWYV pseudoknot inferred from simulations at zero force. Besides folded (F) and unfolded (U) states, there are two intermediate state, I1 and I2. The topology of the I1 state is the same as F except tertiary interactions are absent.

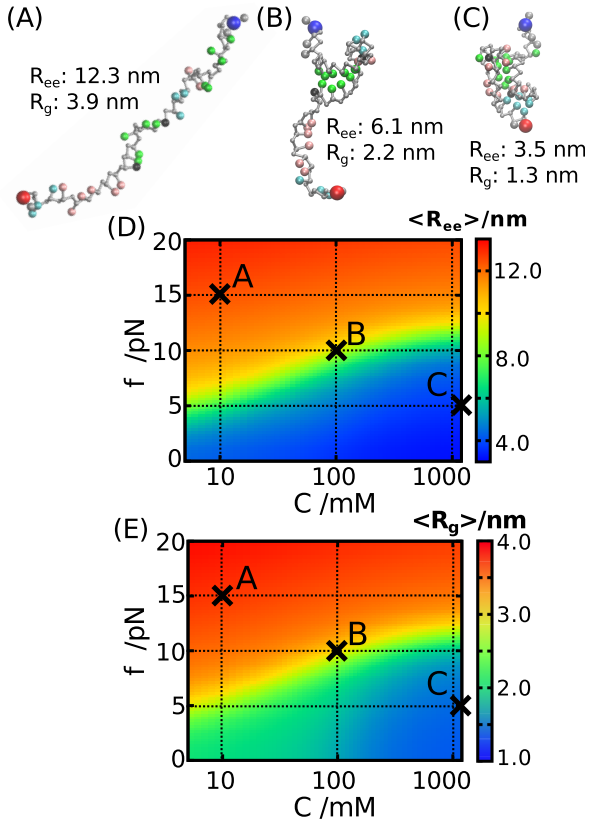


Figure 4: **(A-C)** Representative snapshots corresponding to the three distinct states from simulations. **(D)** Diagram of states in the $[C, f]$ plane obtained using extension R_{ee} as the order parameter. **(E)** Same as D except R_{g} is used to distinguish the states. The three crosses in D and E correspond to conditions from which the snapshots $(A\text{-}C)$ are sampled. The scale for R_{ee} and R_{g} are given on the right.

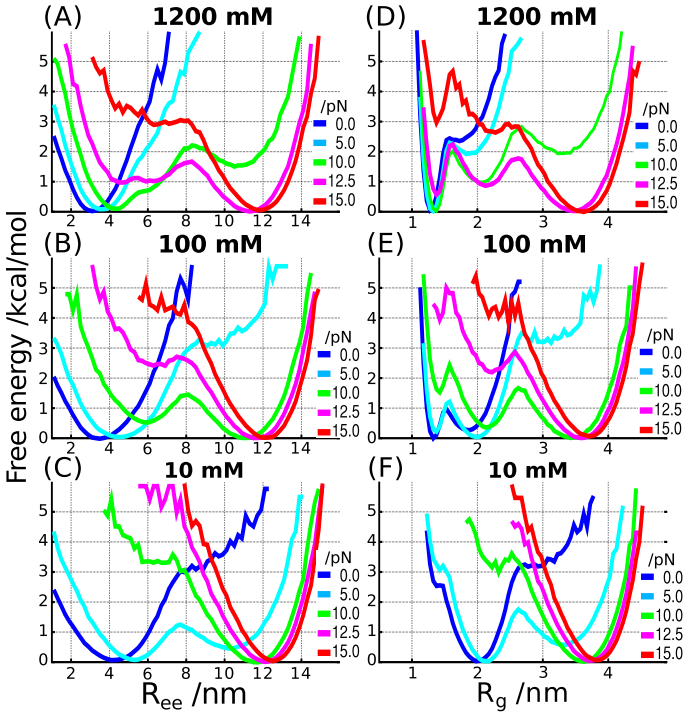


Figure 5: Free energy profiles as a function of R_{ee} **(A-C)** and R_{g} **(D-E)** at three different salt concentrations and a range of forces. The locations of the minima change as C and f vary indicating transitions between distinct states of the BWYV PK.

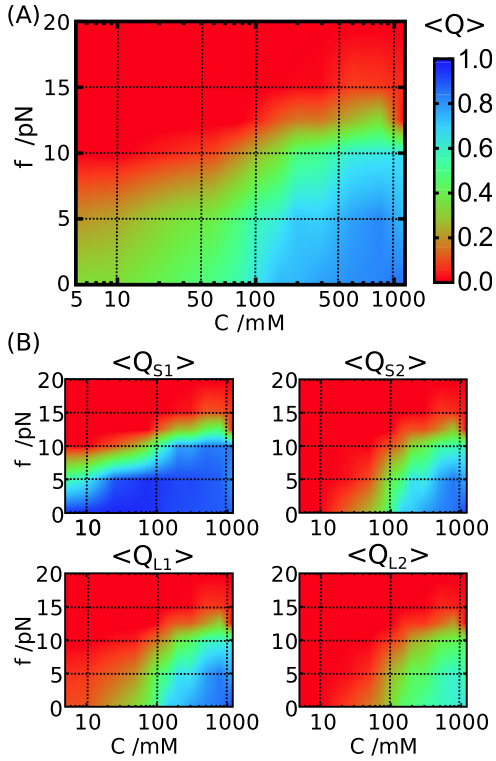


Figure 6: $[C, f]$ phase diagram using the fraction of native contacts as the order parameter. **(A)** The diagram of states determined using an average of the total Q for the PK. The scale is given on the right. **(B)** Decomposition of Q into stems and loops: S1, stem 1; S2, stem 2; L1, loop 1; L2, loop 2.

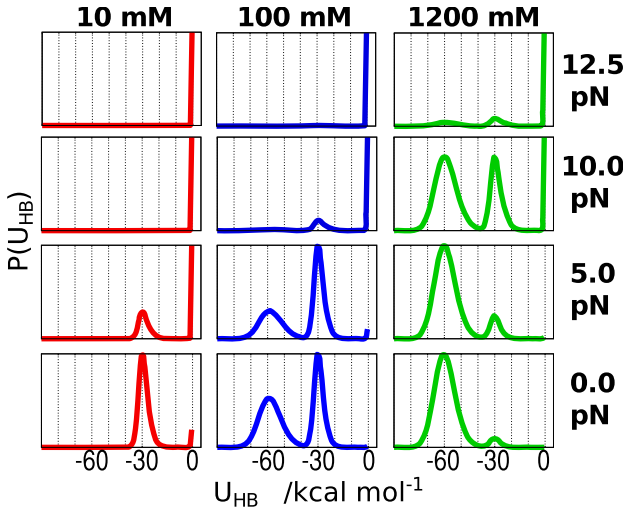


Figure 7: Probability distributions of the total hydrogen bond energy as a function of C and f, which are labeled. The distinct peaks are fingerprints of the states that are populated at a specified C and f.

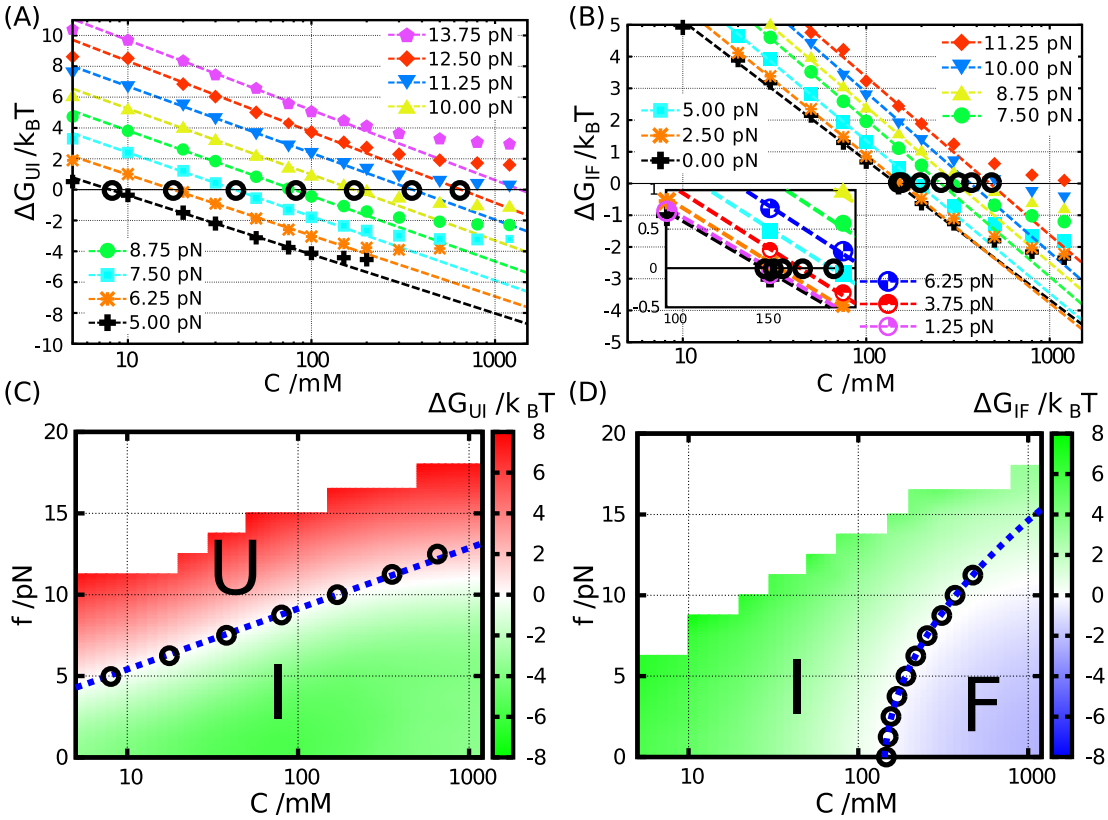


Figure 8: Free energy differences between U and I, and I and F calculated from REMD simulations. The definition of states, U, I, and F, is based on the hydrogen-bond interaction energies associated with each structural element (see main text for details). The I state is analogous to I2 in Figure 3. **(A, B)** Salt concentration dependences of ΔG is plotted for different force values. (Symbols are explained in A and B, respectively.) **(C)** ΔG_{UI} is shown on the $[C, f]$ plane and phase boundary is quantitatively described (white zone between red and green regions). The blue dashed line is the theoretical prediction where $f_{\mathrm{c}} \propto \log C_{\mathrm{m}}$ (Eq.4). Circles ($\circ$) indicate conditions with $\Delta G_{\mathrm{UI}} = 0$ extracted from linear dependence of ΔG on $\log C$ in (A). It should be noted that the U state is under tension, and hence is different from the unfolded state upon thermal melting at $f = 0$. **(D)** Same as (C) except the results describe I$\rightleftharpoons$F transition. Here the critical force describing the phase boundary is a non-linear function of $\log C_{\mathrm{m}}$ (Eq.6).

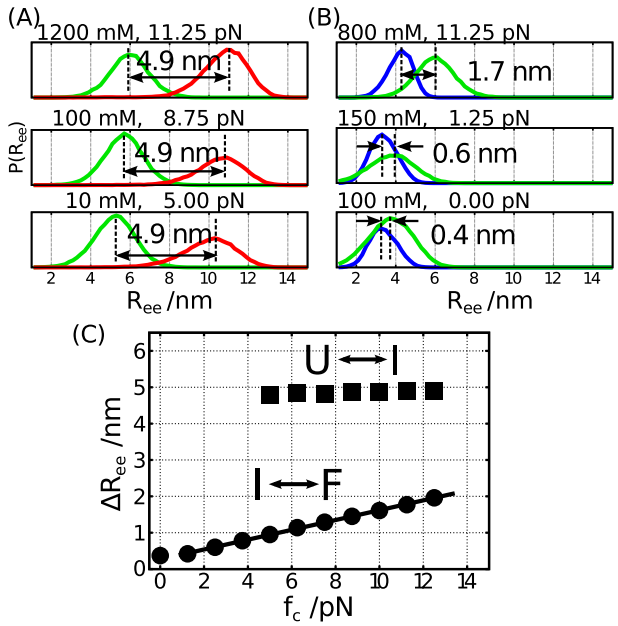


Figure 9: **(A, B)** Probability distributions of the extensions for states U and I (A), and I and F (B). All the conditions shown here are near the phase boundaries in the $[C, f]$ plane (ΔG_{UI} and ΔG_{IF} between the two states are close to 0). Distances between peaks of the two distributions are shown. In the U⇌I transition (A), the mean distances are almost constant around 4.9 nm. On the other hand, in the I⇌F transition (B), the mean distance becomes smaller as the mechanical force weakens. **(C)** Dependence of ΔR_{ee} on f_{c} along the phase boundaries. In the I⇌F transition, the dependence can be fit using $\Delta R_{\mathrm{ee}}^{\mathrm{IF}} = a f_{\mathrm{c}}^{\mathrm{IF}} + b$ where $a = 0.14$ nm/pN and $b = 0.27$ nm (solid line).