Vectorial unfolding of RNA pseudoknot in a protein channel

Xinyue Zhang, Xiaojun Xu, Zhiyu Yang, Andrew J. Burcke, Kent S. Gates, Shi-Jie Chen, and Li-Qun Gu

Supplementary Information

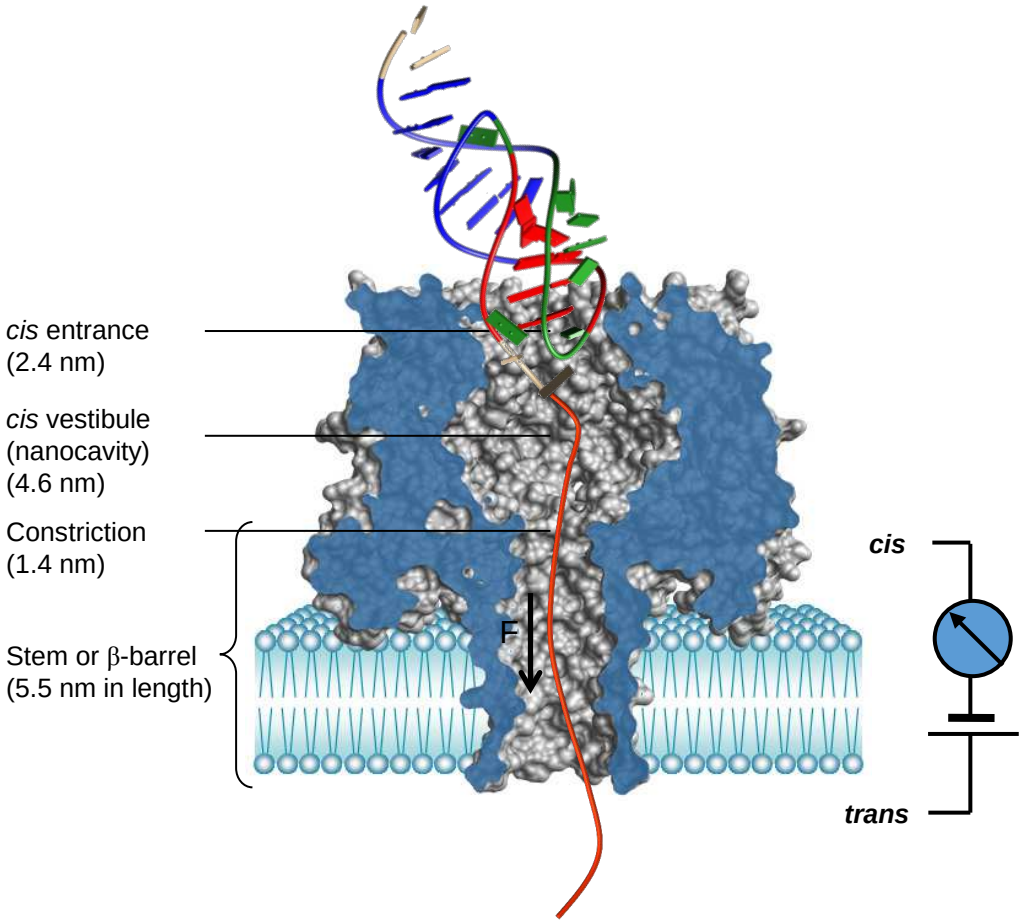


Figure S1. Dimension of the α-hemolysin protein pore (*Science* 274, 1859–1866 (1996)) in the membrane and the electric driving force on the target nucleic acids in the nanopore. The simplified force expression is $F=qNV/L$. In this expression, V is the voltage across the pore. In the pseudoknot pulling experiment, V=+120 mV (*cis* grounded). Most of the voltage drops on the narrow β-barrel with a length of L. L=5.5 nm; N is the number of nucleotides acted by the electric field, and N=11 nucleotides as used in the literature *PNAS* 106, 7702–7707 (2009); q is the effective charge of each nucleotide in the solution. In the literatures, q ranges from -0.1e (Mathe et al *Biophys. J.* 87, 3205–3212 (2004)) to -0.5e (Schellman et al *Biopolymers* 16, 1415–1434 (1977)). We used q=−0.25e in 1 M salt concentration, as reported in Keyser et al *Nat. Phys.* 2, 473-477 (2006). Using these parameter values, we evaluated that F=9.6 pN. The scale of this force is consistent with that reported for driving i-motif unfolding in the nanopore (*J. Am. Chem. Soc.* 137, 9053–9060 (2015)).

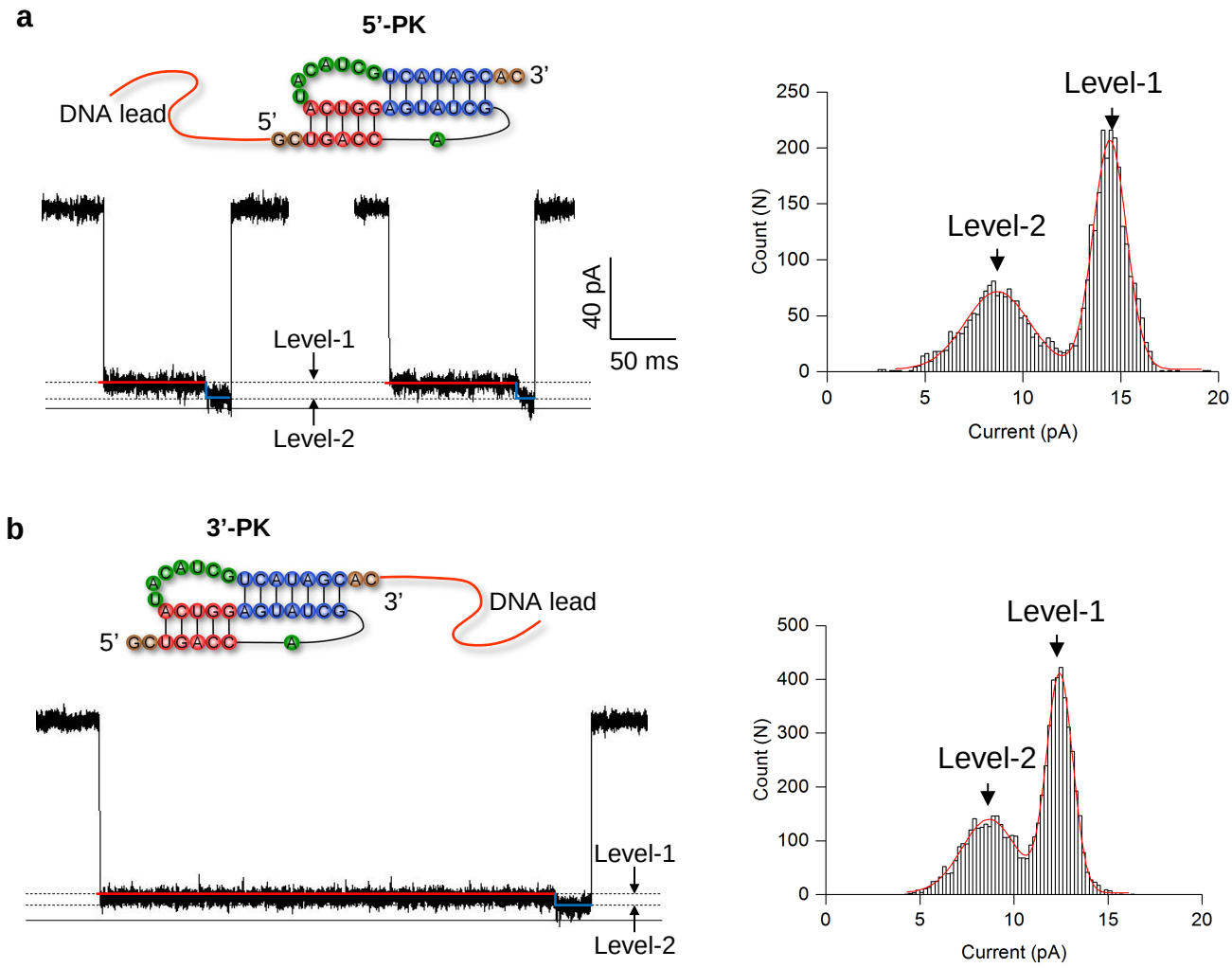


Figure S2. Nanopore blocking events generated by 5'-PK and 3'-PK and conductance analysis. **a)** 5'-PK; **b)** 3'-PK. In both a and b, left panels are current traces, and right panels are corresponding current amplitude histograms, where the peaks for Level-1 and Level-2 in the blocking events are marked. Level-2 is lower than Level-1 for both 5'-PK and 3'-PK.

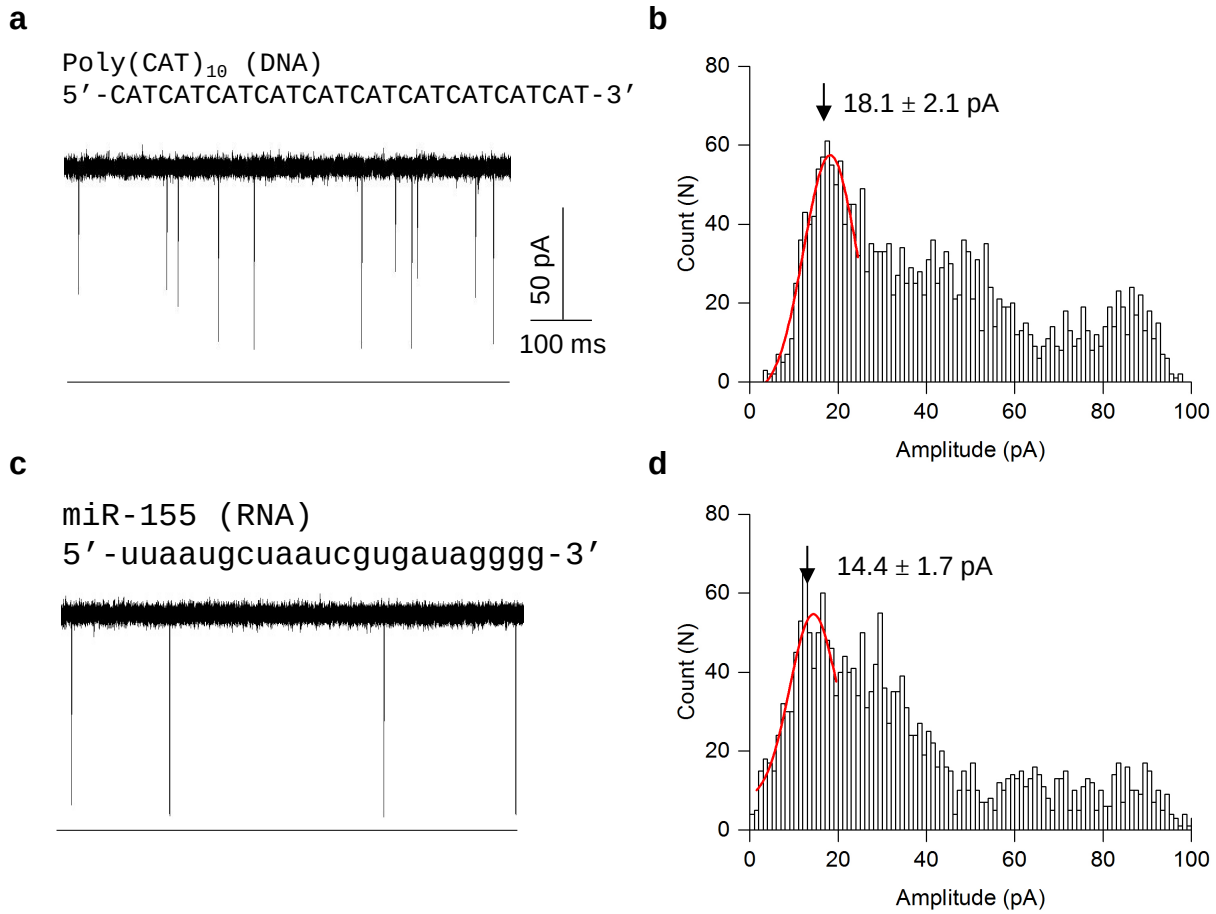


Figure S3. Nanopore blocking events generated by translocation of ssDNA poly(CAT)$_{10}$ and extended RNA miR-155. **a)** poly(CAT)$_{10}$ and **b)** miR-155. In both a and b, left panels are current traces, and right panels are corresponding current amplitude histograms, where the peaks for full blocks by DNA or RNA translocation are marked. Note that RNA translocation reduces more conductance (lower conductance) than DNA translocation events. This comparison helps us to monitor the position of unfolded RNA in the nanopore.

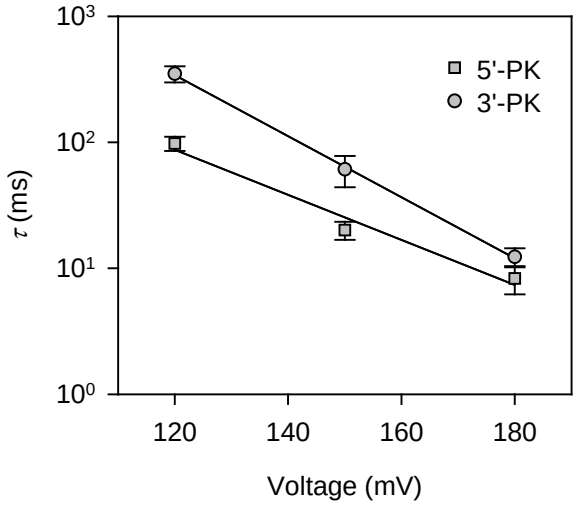


Figure S4. Blocking event duration (τ) of 5'-PK (■) and 3'-PK (●) as a function of transmembrane voltage. For both pseudoknots, τ was shortened as the voltage increases from +120 to +180 mV, verifying the unfolding of T2 pseudoknot pulled by the DNA leader.

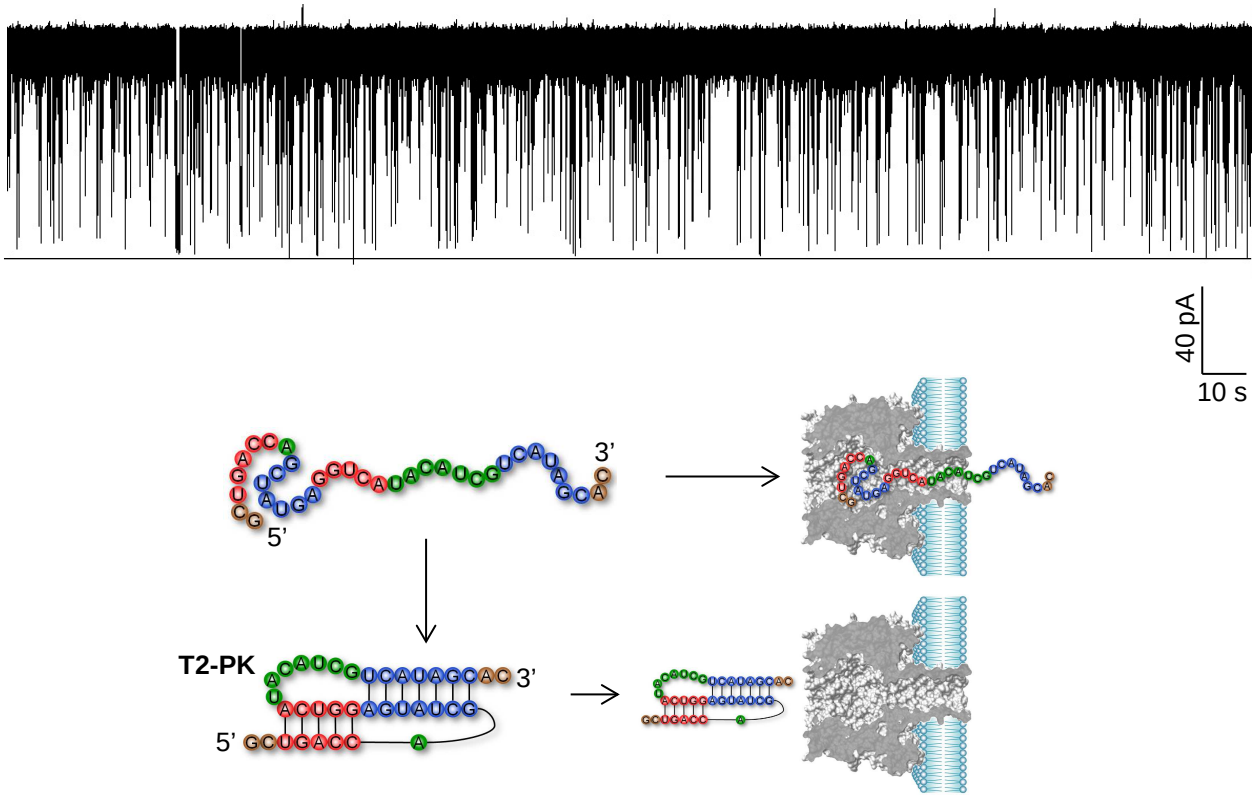
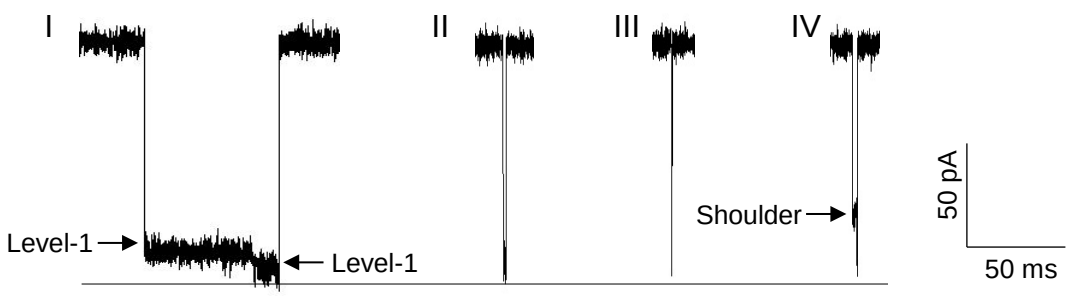


Figure S5. Nanopore blocking events generated by T2-PK RNA (top) and molecular configuration model (bottom). T2-PK is the T2 pseudoknot without tags. Unlike 5'- and 3'-PK, T2-PK (300 nM) only produced short blocks that can be assigned to non-pseudoknot structures such as partially folded hairpins that unzip in the pore and RNAs in the extended forms that translocate through the pore. This result suggests that the pseudoknot molecules formed by T2-PK cannot be trapped in the *cis* vestibule of the pore. This is different from another folded structure, DNA G-quadruplex, which can be trapped in the *cis* vestibule to partially block the nanopore current (I/I_0=50%) [Shim et al *J.Phys.Chem* B 112, 8354-8360 (2008)]. The pseudoknot population for T2-PK was evaluated in Fig. S6.

a



b. Blocking level and duration of various blocks generated chimera (5'-PK)

	Structures	Molecular process	Relative conductance (I/I_0)		Duration (ms)
I	Pseudoknot	Unfolding	15±1% (Level-1)	9.7±0.9% (Level-2)	98±6
II	Hairpins	Unzipping		10±2%	18±3
III	Extended conformation	1-step translocation		9.1±1.4%	0.32±0.20
IV	Extended conformation	2-step translocation	37±4.0% (Shoulder)	9.7±0.8% (Ending spike)	1.7±0.2

c. Blocking level and duration of various blocks generated chimera (3'-PK)

	Structures	Molecular process	Relative conductance (I/I_0)		Duration (ms)
I	Pseudoknot	Unfolding	13±1% (Level-1)	9.2±1.5% (Level-2)	350±40
II	Hairpins	Unzipping		9.1±1.2%	23±4
III	Extended conformation	1-step translocation		8.8±2.0%	0.41±0.18
IV	Extended conformation	2-step translocation	38±5% (Shoulder)	8.1±1.1% (Ending spike)	2.5±0.7

d. Populations of different types of blocks generated by T2 pseudoknot and variants

	Structures	T2-PK [a]	5'-PK	3'-PK	5'-PK-mut	3'-PK-mut
I	Pseudoknot	40%	56%	51%	12%	16%
II	Hairpins	-	11%	13%	21%	26%
III	Extended conformation	-	27%	29%	45%	39%
IV	Extended conformation	-	6%	7%	22%	19%

[a]: T2-PK is a pseudoknot without tags. But its blocking events were not observed, suggesting that pseudoknot alone cannot access the pore. Therefore, the pseudoknot population of T2-PK was evaluated by comparing the non-pseudoknot structure capture rate between T2-PK and tagged pseudoknot (5'-PK). For example, for 3'-PK, the frequency of all blocks, including pseudoknot and non-pheudoknot, was 0.60 s^{-1} in 100 nM chimera concentration. The overall capture rate was calculated to be 6.0 $\mu M^{-1} \cdot s^{-1}$. From the block breakdown table (panel d), we have known that pseudoknot population was 51%. So the capture rate contributed by all non-pseudoknot molecules would be 3.0 $\mu M^{-1} \cdot s^{-1}$. For T2-PK, all blocks we observed were non-pseudoknot signatures (for hairpins unzipping, and 1- and 2-step translocations). Their frequency was 1.1 s^{-1} in 300 nM T2-PK RNA. So the contribution of non-pseudoknot to the overall capture rate was 3.6 $\mu M^{-1} \cdot s^{-1}$. Assuming the overall capture is 6.0 $\mu M^{-1} \cdot s^{-1}$, the non-pseudoknot population should be 61%. Therefore the pseudoknot population of T2-PK was evaluated to be 40%.

Figure S6. a) Representative block patterns generated by chimera (5'-PK). The chimera can form pseudoknot that unfolds and translocates in the nanopore (I) and partially folded hairpins that are unzipped and translocate in the pore (II) (predicted using OligoAnalyzer 3.1, Integrated DNA Technologies, Inc). The chomeras can also adopt extended conformation for rapid translocation through the pore in one step (III) and two step (IV). **b-c)** Properties of observed block patterns generated by 5'-PK (b) and 3'-PK (c). **d)** Population of various block patterns generated by T2 pseudoknot and its variants, including T2-PK, 5'-PK, 3'-PK, 5'-PK'-mut, and 3'-PK-mut.

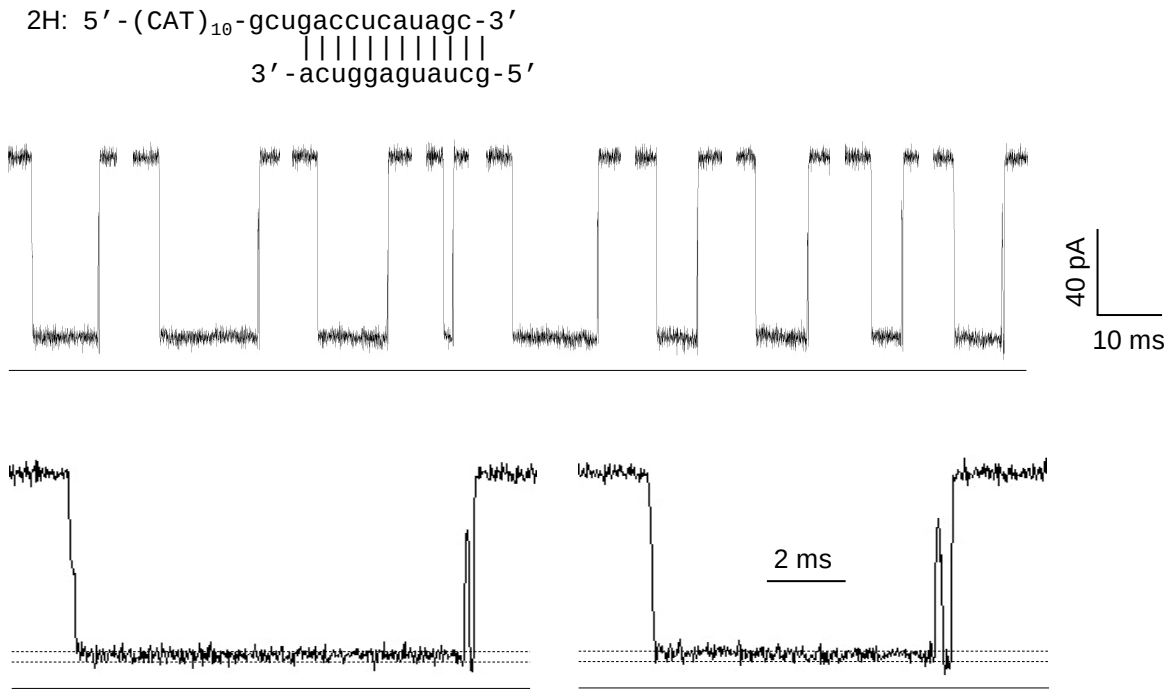


Figure S7. Blocking events for unzipping and translocation of a double stranded RNA 2H. This 12-bp RNA duplex consists of the 7-bp H1 and 5-bp H2 without loop participation (top). A poly(CAT)$_{10}$ DNA leader was attached as an overhang to trap the RNA duplex into the pore and drive its unfolding. The current traces for 2H (mid) and expanded events (bottom) show that the unzipping of 2H was only tens of milliseconds, much shorter than the unfolding of 5'-PK and 3'-PK that involves the non-canonical interactions such as triplex base-pairs for structural stabilization.

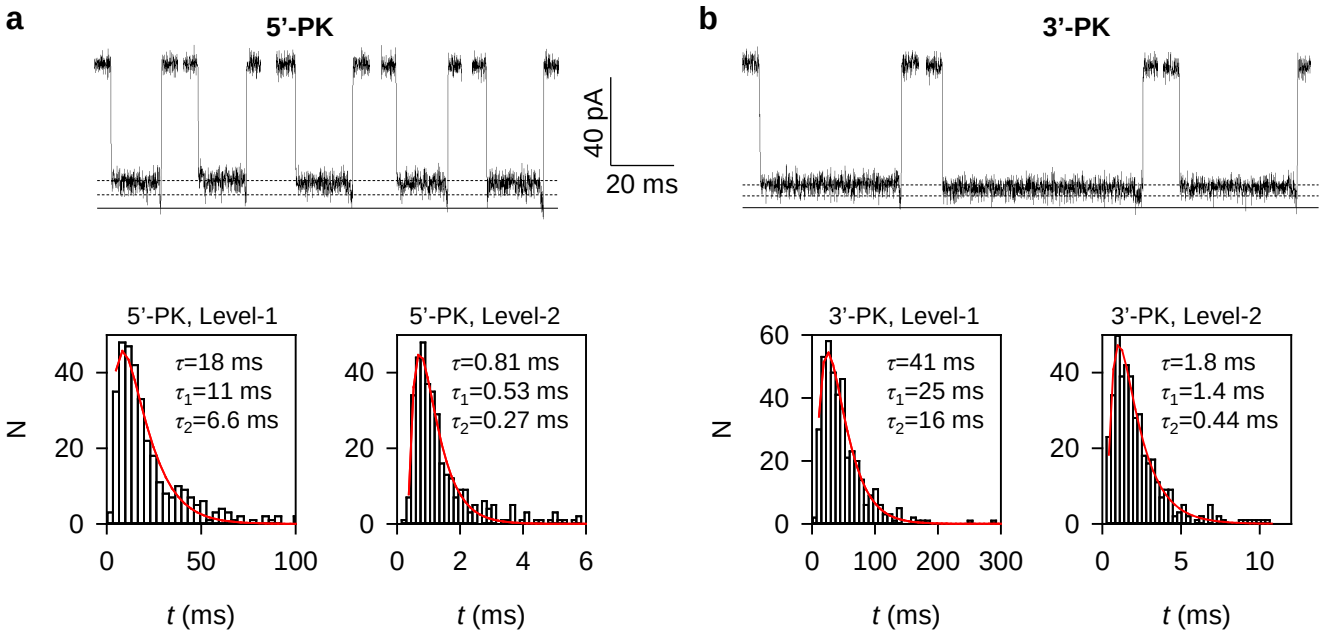


Figure S8. Nanopore blocking events of 5'-PK and 3'-PK and duration of different steps in signature blocks in near physiology condition. Current traces were recorded at +120 mV in asymmetrical solutions: *cis* (5'-PK or 3'-PK presented), 140 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$; *trans*, 2 M NaCl. Both *cis* and *trans* solutions were buffered with 25 mM MOPS (pH 7.4) and in the presence of 10 mM $MgCl_2$. **a)** and **b)** Blocking events (top) and representative Level-1 and Level-2 duration histograms (bottom) for 5'-PK (N=361, panel a) and 3'-PK (N=410, panel b). The histograms were fitted with $N(t) = \frac{A}{\tau_1 - \tau_2}\left[e^{-t/\tau_1} - e^{-t/\tau_2}\right]$, where τ_1 and τ_2 are the two sequential time components. The duration of each blocking level, τ, is the sum of τ_1 and τ_2.

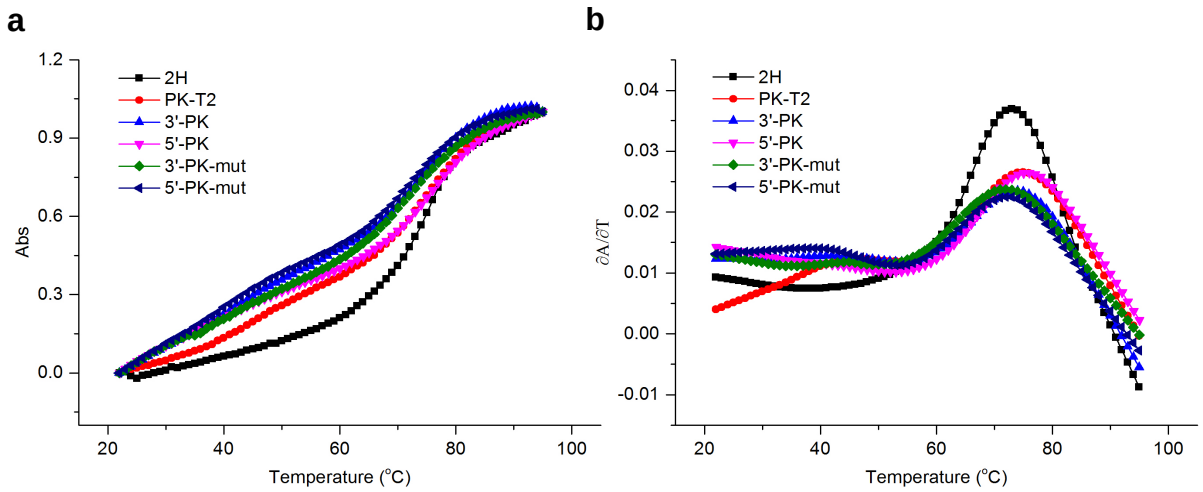


Figure S9. Thermal stability of T2 pseudoknot and its regulation by loop-stem interactions. **a)** Melting curves of 5'-PK, 3'-PK, 5'-PK-mut, 3'-PK-mut, T2-PK, and 2H. **b)** First derivative of melting curves in panel a, as processed in the reference [Qin et al *Biochemistry* 35, 4176-4186 (1996)]. All nucleic acids possessed a derivative peak around 74-76 °C, which is the melting temperature of coaxial-stacking helix. Except 2H, all other pseudoknot-forming sequences show another derivative peak around 40-45 °C, which could be attributed to the disruption of loop-helix triplex interactions.

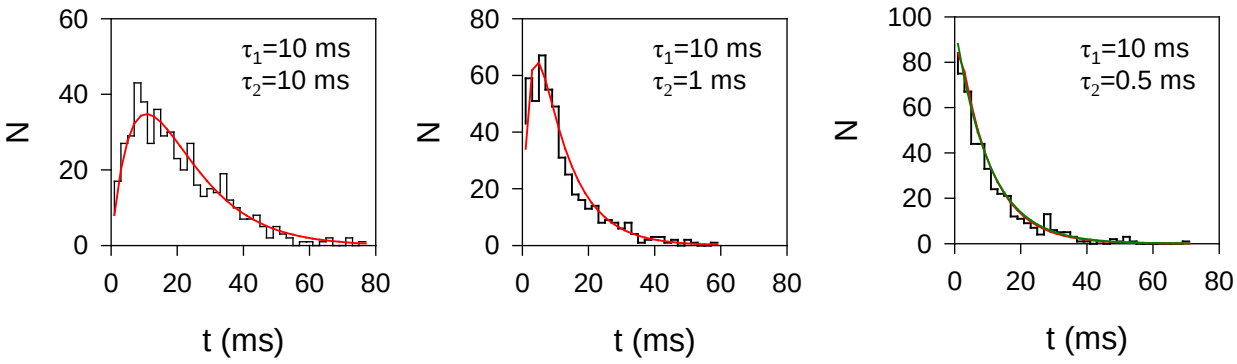


Figure S10. Simulation of the two-step kinetics and the fitting of time constants. In the A→B→C kinetic pathway, the lifetimes of a molecule in A and B states are exponentially distributed with the constants of τ_1 and τ_2. The total time of A and B states τ can be fitted with a two-exponential distribution $N(t) = \frac{A}{\tau_1 - \tau_2}\left[e^{-t/\tau_1} - e^{-t/\tau_2}\right]$. τ is the sum of τ_1 and τ_2. Compared with the single-exponential distribution, such two-exponential distribution is characterized by a peak, which is the key to the fitting accuracy. To simulate the fitting procedure, we generated two sets of 200 exponentially-distributed duration values for A and B, respectively, to simulate the fitting process and to study the condition for accurate fitting. **a)** When τ_1/τ_2 =1 (τ_1=10 ms, τ_1=10 ms), the peak time is the longest, and the fitting is the most accurate (τ_1=10.7 ms, τ_1=10.4 ms) with lowest relative errors (τ_1: 7% and τ_2: 4%). **b)** When τ_1/τ_2 =5 (τ_1=10 ms, τ_1=2 ms), the peak time is shortened, and the fitting, in particular the shorter time τ_2 is less accurate (τ_1=9.6 ms, τ_2=2.3 ms) with increased errors (τ_1: 4% and τ_2: 15%). **c)** When τ_1/τ_2 =20 (τ_1=10 ms, τ_1=0.5 ms), the peak time is greatly shortened (τ_1=9.7 ms, τ_2=0.37 ms) with the increase in the errors (τ_1: 3% and τ_2: 23%). Note that in this case, the smaller time componentco τ_2 can be negected and the histogram can be fitted with a single-exponent distribution. Therefore, empiracally, if the peak time in a histogram is not zero, the distribution can be fitted with a two exponential curve. In contrast, if the peak time is zero (i.e. the first bin is the largest), we can treat the histogram as the single-exponent distribuion.

Table S1 Energy parameters used for the KMC simulations

Energy parameters	References
Base (mismatched) stacking	*Nucleic Acids Res*. (2010), 38: D280-282
Hairpin loops	*RNA* (2005), 11: 1884-1897
Internal loops	*RNA* (2005), 11: 1884-1897
Bulge loops	*RNA* (2005), 11: 1884-1897
Pseudoknots without inter-helix loop	*Nucleic Acids Res*. (2006), 34: 2634-2652,
Pseudoknots with inter-helix loop	*RNA* (2009), 15: 696-706
Base triples	*RNA* (2010), 16: 538-552