Published in final edited form as:

J Am Chem Soc. 2018 August 15; 140(32): 10067–10070. doi:10.1021/jacs.8b03784.

Mimicking Co-Transcriptional RNA Folding Using a Superhelicase

Boyang Hua[#†], **Subrata Panja**[##], **Yanbo Wang**[†], **Sarah A. Woodson**[*,#], and **Taekjip Ha**[*,†,#,^,§]

[†]Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, United States

[#]T. C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, United States

[^]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States

[§]Howard Hughes Medical Institute, Baltimore, Maryland 21205, United States

[#] These authors contributed equally to this work.

Abstract

Vectorial folding of RNA during transcription can produce intermediates with distinct biochemical activities. Here, we design an artificial minimal system to mimic co-transcriptional RNA folding *in vitro*. In this system, a pre-synthesized RNA molecule begins to fold from its 5'-end, as it is released from a heteroduplex by an engineered helicase that translocates on the complementary DNA strand in the 3'-to-5' direction. This chemically stabilized, "superhelicase" Rep-X processively unwinds thousands of base pairs of DNA. The pre-synthesized RNA enables us to flexibly position fluorescent labels on the RNA for single-molecule fluorescence resonance energy transfer analysis and allows us to study real-time conformational dynamics during the vectorial folding process. We observed distinct signatures of the maiden secondary and tertiary folding of the *Oryza sativa* twister ribozyme. The maiden vectorial tertiary folding transitions occurred faster than Mg^{2+}-induced refolding, but were also more prone to misfolding, likely due to sequential formation of alternative secondary structures. This novel assay can be applied to studying other kinetically controlled processes, such as riboswitch control and RNA–protein assembly.

Graphical abstract

[*]**Corresponding Author**: swoodson@jhu.edu (S. A. W.) and tjha@jhu.edu (T. H.).

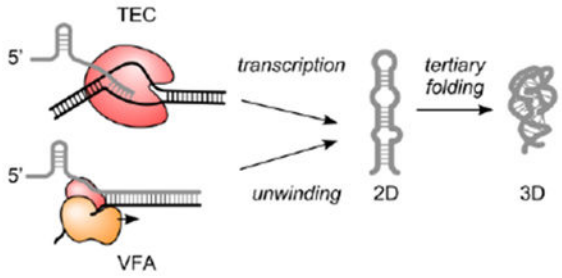
ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b03784.

Experimental details and supplementary figures (PDF)

The authors declare no competing financial interest.



Accurate folding is essential for many RNA functions. Because the 5'-end of an RNA molecule begins to fold before the 3'-end is synthesized, co-transcriptional folding can produce intermediates with biochemical activities that are distinct from refolded RNA *in vitro* (Figure 1a top).[1–9]

Co-transcriptional RNA folding has been studied by various methods.[10, 11] For example, chemical footprinting was combined with next-generation sequencing techniques to map the co-transcriptional RNA folding pathway at nucleotide resolution *in vitro* and *in vivo*.[12, 13] Unlike structure probing methods, single-molecule methods can resolve different folding pathways within an RNA population.[14] An optical-trapping assay was developed to study co-transcriptional folding of the *pbuE* adenine riboswitch.[15] Recently, a single-molecule fluorescence resonance energy transfer (FRET) assay was developed to study the *thiM* thiamine pyrophosphate riboswitch.[16]

Here, we describe a superhelicase-based method that mimics the vectorial nature of co-transcriptional RNA folding with real-time single-molecule observation (Figure 1a bottom). In this vectorial folding assay (VFA in Methods), a fluorescently labeled RNA is synthesized and annealed to its complementary DNA (cDNA) to form a DNA-RNA heteroduplex (Figure S1a,b). A highly processive superhelicase Rep-X, engineered from *E. coli* Rep helicase through chemical crosslinking to stabilize the unwinding active conformation,[17, 18] is loaded on the heteroduplex in the absence of $MgCl_2$ and ATP. After incubation, free Rep-X is washed out and an unwinding buffer containing $MgCl_2$ and ATP is added to start duplex unwinding. The RNA strand begins to fold as the bound helicase translocates 3'-to-5' along the cDNA strand, unwinding the duplex and releasing the RNA strand in the direction of transcription. We detect real-time folding events of single surface-tethered molecules through FRET between strategically positioned fluorophores.

Although this pseudo-co-transcriptional folding scheme cannot recapitulate specific interactions with elongation factors[8] or sequence-dependent pausing,[4, 19] an important aspect of co-transcriptional RNA folding, the vectorial directionality can be mimicked. In this study, we used the naturally occurring *Oryza sativa* (*Osa*) twister ribozyme (Figure 1b and ref. 20) to demonstrate how this assay can reveal differences between refolding and vectorial folding of pre-synthesized RNA.

Nearly the entire population of molecules that have fluorescently active donor (Cy3) and acceptor (Cy5) fluorophores stayed in the heteroduplex before unwinding [FRET efficiency (E_{FRET}) ~ 0.05, Figure 1c top]. After loading Rep-X (20 nM) and subsequent addition of 20

mM $MgCl_2$ and 1 mM ATP, about one third of the doubly labeled molecules displayed E_{FRET} transitions between ~0.2 and ~0.9 (Figure 1c middle and Figure S1c), corresponding to transitions between the 2D structure and folded 3D structure.[20] Although higher than physiological concentrations, we chose 20 mM $MgCl_2$, the 3D folding midpoint of *Osa* twister[20], because frequent folding and unfolding transitions facilitate the search for the native structure.[21] The folding and unfolding rates at equilibrium were similar whether the RNA was folded vectorially (VFA) or refolded by adding $MgCl_2$ (RFA in Methods, Figure 1c bottom and Figure S1d,e), suggesting that the helicase and cDNA strand were released and did not alter the folding properties of the tethered twister molecules.

The normalized unwinding yield was 35% (Figure 1d, unwinding yield normalization in Methods). Controls lacking either Rep-X or ATP showed <3% yields (Figure 1d). This baseline yield is likely due to residual unhybridized RNA. Replacing the cross-linked Rep-X with an uncrosslinked double-cysteine mutant Rep (RepDM) greatly reduced the normalized unwinding yield (<10%, Figure 1d), consistent with the previous finding that cross-linking increases the Rep helicase activity by stabilizing the unwinding active conformation.[17, 18] We determined the dissociation constant between the heteroduplex and Rep-X to be 9.2 nM during pre-incubation (Figure S2). We also demonstrated that the superhelicase-based VFA works at a wide range of salt and ATP concentrations (Figure S2).[22] Optimized con ditions for helicase loading and unwinding were used for the rest of the study (Methods).

From real-time single-molecule fluorescence signals, we observed an intensity peak in the donor signal after adding the unwinding buffer, which we interpreted as protein-induced fluorescence enhancement (PIFE) when the helicase translocated past the Cy3 fluorophore attached to U24 (Figure 2a blue arrow and Figure S3).[23, 24] Following the PIFE peak, we observed a transient spike in E_{FRET} (Figure 2a green arrow) through which the E_{FRET} increased from ~ 0.05 to ~ 0.2. We interpreted this spike as coiling of the single-stranded RNA (higher E_{FRET} 0.4–0.6), [17, 25] followed by the maiden 2D folding transition (E_{FRET} ~ 0.2). The subsequent maiden 3D folding transition resulted in an E_{FRET} increase from ~ 0.2 to > 0.3, which includes native and non-native 3D structures (Figure 2a magenta arrow).

Trajectories showing maiden 2D or 3D transitions, indicating successful unwinding, were further analyzed. Among the trajectories with successful unwinding, ~ 30% showed all three signatures (the PIFE peak and the maiden 2D and 3D folding transitions). The possible reasons for missing one or two signatures in the remaining trajectories include the limited time resolution and signal-to-noise ratio of our measurements, as well as premature photobleaching. Similar changes in PIFE and FRET were observed when Rep-X was replaced by another superhelicase PcrA-X with a slower unwinding speed (Figure S4).[17]

We synchronized multiple “flow-in” movies to offset the difference in their ATP injection times (Methods) and plotted the times when the PIFE peaks and the maiden 2D folding transitions occurred, respectively (Figure 2c). The time delay between the average PIFE peak center and the average maiden 2D folding time ($\Delta t_{2D–U24} = \langle t_{2D} \rangle - \langle t_{U24} \rangle$) was 1.4 s. Given that 2D structure typically forms in tens of microseconds to milliseconds,[1, 2] most of $\Delta t_{2D–U24}$ was spent by Rep-X unwinding the last 30 bp of the heteroduplex (from U24 to

G54). Therefore, the estimated unwinding speed of Rep-X was ~ 21 bp/s, close to the lower limit measured on double-stranded DNA.[17]

In VFA, the maiden 3D folding transitions occurred faster than in RFA (Figure 3a). 40% (21/53) of all maiden 3D folding transitions in VFA occurred within 1 s after the maiden 2D folding transitions, while only 14% (8/57) of all maiden 3D refolding transitions occurred within 1 s after $MgCl_2$ addition. Most of the "fast" maiden 3D folding transitions in VFA formed the native conformation as reported through $E_{FRET} > 0.75$ [86% (18/21)], consistent with theoretical predictions of folding free energy landscapes.[1]

Overall, 25% (13/53) of the released twister molecules misfolded in their first attempts with a E_{FRET} value lower than the native conformation (E_{FRET} 0.3 to 0.75, Figure 3b,e and Figure S5). Most of the initially misfolded molecules recovered at equilibrium; only 12% (13/112) stayed misfolded after 5 min (Figure 3c,e and Figure S5). The maiden 3D folding transition in VFA was more heterogeneous than later folding transitions, as twister often misfolded first and subsequently unfolded and formed its native structure.

To further investigate the cause of heterogeneity in the maiden 3D folding, we performed RFA of twister molecules. In this assay, the 2D structures of twister were fully equilibrated before we initiated refolding by supplying $MgCl_2$ (Methods). Only 16% (9/57) of twister molecules misfolded in their first attempts and 7% (8/116) remained misfolded at equilibrium (Figure 3b,c,e and Figure S5). The higher propensity for maiden 3D misfolding in VFA suggested that some of the maiden 2D structures were incorrect, leading to non-native 3D structures, although we cannot exclude the potential interference of folding by the single-stranded cDNA.

To directly probe the maiden formation of 2D structure in VFA, we relocated the Cy3 fluorophore to the 5'-end of the RNA (C1) (Figure S6). We observed a large E_{FRET} increase from ~ 0.05 to ~ 1.0 following the PIFE peak, representing the maiden 2D folding transitions. The time delay between the average PIFE peak center and the average maiden 2D folding time ($\Delta t_{2D-C1} = \langle t_{2D} \rangle - \langle t_{C1} \rangle$) was 2.3 s (Figure S6), yielding an unwinding speed (~ 24 bp/s) close to that obtained with the Cy3 fluorophore at U24. Most of the released twister strands could form a stable 2D structure (E_{FRET} ~ 1.0, Figure S6), although 40% (17/43) of them went through misfolded intermediates ($E_{FRET} < 0.5$, Figure S7). These results suggested that the freshly released twister strands could access non-native base-paired structures.

Simulation of the co-transcriptional folding pathway of twister revealed a plausible non-native 2D structure where a subtle 3-bp shift in the P1 stem blocked the two guanine bases (G47 and G48) essential for forming pseudoknot 2 (PK2, Figure 3f), suggesting the PK2 perturbation could be a source of the observed maiden folding heterogeneity. However, our labeling strategy is insensitive to probe the small distance change indicated by the simulation. Therefore, we mutated G48 to A to directly perturb PK2. Equilibrium RFA with the Cy3 fluorophore at U24 showed that G48A mutant RNA formed 3D structures only transiently, and often formed 3D structures with $E_{FRET} < 0.7$ [40% (28/70), Figure 3d,e and Figure S8], similar to the misfolded population of the wild-type twister during maiden

folding in VFA. Instability of the P1 stem was also found to cause misfolding of the *env-22* twister ribozyme, further suggesting P1 as a likely source of folding heterogeneity.[26]

In summary, we utilized chemically engineered superhelicases to mimic co-transcriptional RNA folding *in vitro*. We observed distinct signatures of 2D and 3D structure during the maiden round of folding in real-time. Vectorial folding differs most from refolding when the lifetimes of 5' RNA structures exceed the elongation time. We find that this can be achieved even with the 54-nt twister ribozyme. The maiden 3D folding transitions in VFA occurred faster than in RFA. VFA allows some twister molecules to access misfolded maiden 2D structures, and these misfolded 2D structures lead to a higher propensity for 3D misfolding in VFA. This novel assay can be applied to studying other kinetically controlled processes, such as riboswitch control and RNA–protein assembly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Aaron L. Lucius for helpful comments in manuscript preparation. This work was supported by U.S. NSF grant PHY 1430124 and NIH grant GM 112659 to T.H. and NSF grant MCB 1616081 to S.A.W. T.H. is an investigator of the Howard Hughes Medical Institute.

REFERENCES

(1). Thirumalai D, Woodson SA (1996) Kinetics of Folding of Proteins and RNA. Acc. Chem. Res. 29, 433–439.

(2). Pan T, Sosnick T (2006) RNA folding during transcription. Annu. Rev. Biophys. Biomol. Struct. 35, 161–175. [PubMed: 16689632]

3). Lai D, Proctor JR, Meyer IM (2013) On the importance of co-transcriptional RNA structure formation. RNA 19, 1461–1473. [PubMed: 24131802]

(4). Wickiser JK, Winkler WC, Breaker RR, Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol. Cell 18, 49–60. [PubMed: 15808508]

(5). Poot RA, Tsareva NV, Boni IV, van Duin J (1997) RNA folding kinetics regulates translation of phage MS2 maturation gene. Proc. Natl. Acad. Sci. U. S.A. 94, 10110–10115. [PubMed: 9294171]

6). Hollands K, Sevostiyanova A, Groisman EA (2014) Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. Proc. Natl. Acad. Sci. U. S. A 111, E1999–2007. [PubMed: 24778260]

(7). Mooney RA, Artsimovitch I, Landick R (1998) Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. J. Bacteriol. 180, 3265–3275. [PubMed: 9642176]

(8). Pan T, Artsimovitch I, Fang XW, Landick R, Sosnick TR (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. Proc. Natl. Acad. Sci. U. S. A 96, 9545–9550. [PubMed: 10449729]

(9). Koduvayur SP, Woodson SA (2004) Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice. RNA 10, 1526–1532. [PubMed: 15337845]

(10). Heilman-Miller SL, Woodson SA (2003) Perturbed folding kinetics of circularly permuted RNAs with altered topology. J. Mol. Biol. 328, 385–394. [PubMed: 12691747]

(11). Perdrizet GA, .Artsimovitch I, Furman R, Sosnick TR, Pan T(2012) Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. Proc. Natl. Acad. Sci. U.S. A. 109, 3323–3328. [PubMed: 22331895]

(12). Watters KE, Strobel EJ, Yu AM, Lis JT, Lucks JB (2016) Cotranscriptional folding of a riboswitch at nucleotide resolution. Nat. Struct. Mol. Biol 23, 1124–1131. [PubMed: 27798597]

(13). Incamato D, Morandi E, Anselmi F, Simon LM, Basile G, Oliviero S (2017) In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. Nucleic Acids Res. 45, 9716–9725. [PubMed: 28934475]

(14). Ray S, Widom JR, Walter NG (2018) Life under the Microscope: Single-Molecule Fluorescence Highlights the RNA World. Chem. Rev 118,4120–4155. [PubMed: 29363314]

(15). Frieda KL, Block SM (2012) Direct observation of cotranscriptional folding in an adenine riboswitch. Science 338, 397–400. [PubMed: 23087247]

(16). Uhm H, Kang W, Ha KS, Kang C, Hohng S (2018) Single-molecule FRET studies on the cotranscriptional folding of a thiamine pyrophosphate riboswitch. Proc. Natl. Acad. Sci. U. S. A. 115, 331–336. [PubMed: 29279370]

(17) . Arslan S, Khafizov R, Thomas CD, Chemla YR, Ha T (2015) Engineering of a superhelicase through conformational control. Science 348, 344–347. [PubMed: 25883358]

(18). Comstock MJ, Whitley KD, Jia H, Sokoloski J, Lohman TM, Ha T, Chemla YR (2015) Direct observation of structure-function relationship in a nucleic acid processing enzyme. Science 348, 352–354. [PubMed: 25883359]

(19). Wong TN, Sosnick TR, Pan T (2007) Folding of noncoding RNAs during transcription facilitated by pausing-induced normative structures. Proc. Natl. Acad. Sci. U. S. A. 104, 17995–18000. [PubMed: 17986617]

(20). Panja S, Hua B, Zegarra D, Ha T, Woodson SA (2017) Metals induce transient folding and activation of the twister ribozyme. Nat. Chem. Biol 13, 1109–1114. [PubMed: 28825710]

(21). Camacho CJ, Thirumalai D (1993) Kinetics and thermodynamics of folding in model proteins. Proc. Natl. Acad. Sci. U. S. A. 90, 6369–6372. [PubMed: 8327519]

(22). Lohman TM, Chao K, Green JM, Sage S, Runyon GT (1989) Large-scale purification and characterization of the Escherichia coli rep gene product. J. Biol. Chem. 264, 10139–10147. [PubMed: 2524489]

(23). Fischer CJ, Maluf NK, Lohman TM (2004) Mechanism of ATP-dependent translocation of E. coli UvrD monomers along single-stranded DNA. J. Mol. Biol. 344, 1287–1309. [PubMed: 15561144]

(24). Hwang H, Kim H, Myong S (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. Proc. Natl. Acad. Sci. U. S. A 108, 7414–7418. [PubMed: 21502529]

(25). Lee G, Bratkowski MA, Ding F, Ke A, Ha T (2012) Elastic coupling between RNA degradation and unwinding by an exoribonuclease. Science 336, 1726–1729. [PubMed: 22745434]

(26). Vusurovic N, Altman RB, Terry DS, Micura R, Blanchard SC (2017) Pseudoknot Formation Seeds the Twister Ribozyme Cleavage Reaction Coordinate. J. Am. Chem. Soc. 139, 8186–8193. [PubMed: 28598157]

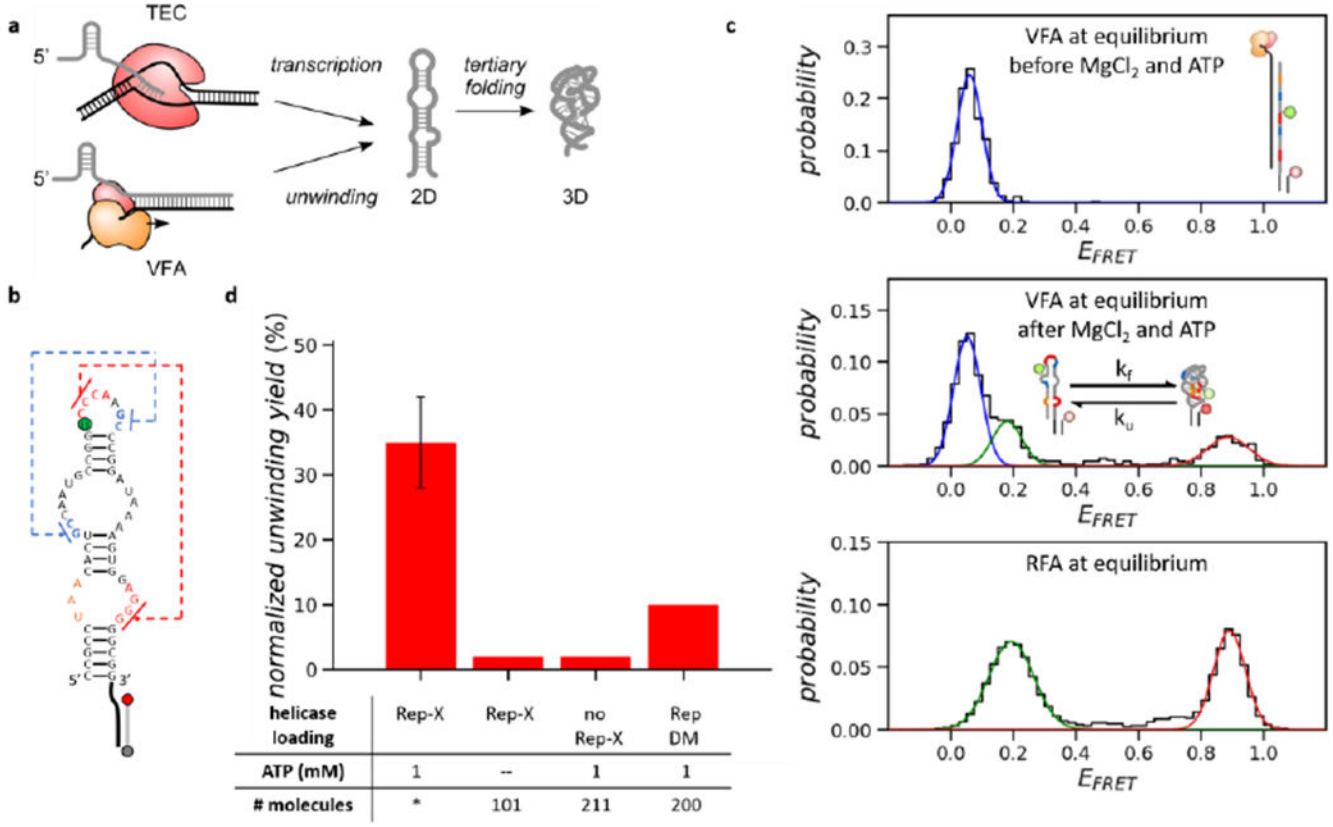


Figure 1.

Characterization of VFA at equilibrium, (a) Schematics of co-transcriptional vs. vectorial folding. TEC: transcription elongation complex. (b) 2D structure of the *Osa* twister ribozyme (length 54 nt). Two pseudoknots are indicated by the blue and red dashed lines. U24 (green) is labeled with the Cy3 fluorophore. The 3'- and 5'-end of the tether strand hg labeled with the Cy5 fluorophore (red) and biotin (grey), respectively, (c) E_{FRET} histograms of twister molecules before (top) and after (middle) the unwinding buffer addition in VFA and in RFA (bottom). Both assays were performed at 20 mM $MgCl_2$. (d) The normalized unwinding yields of VFA. The helicase loading concentration was 20 nM (Rep-X or RepDM). During unwinding, 20 mM $MgCl_2$ was used in all conditions of VFA. * Independent repeats (n = 9) and the error bar represents the standard deviation.

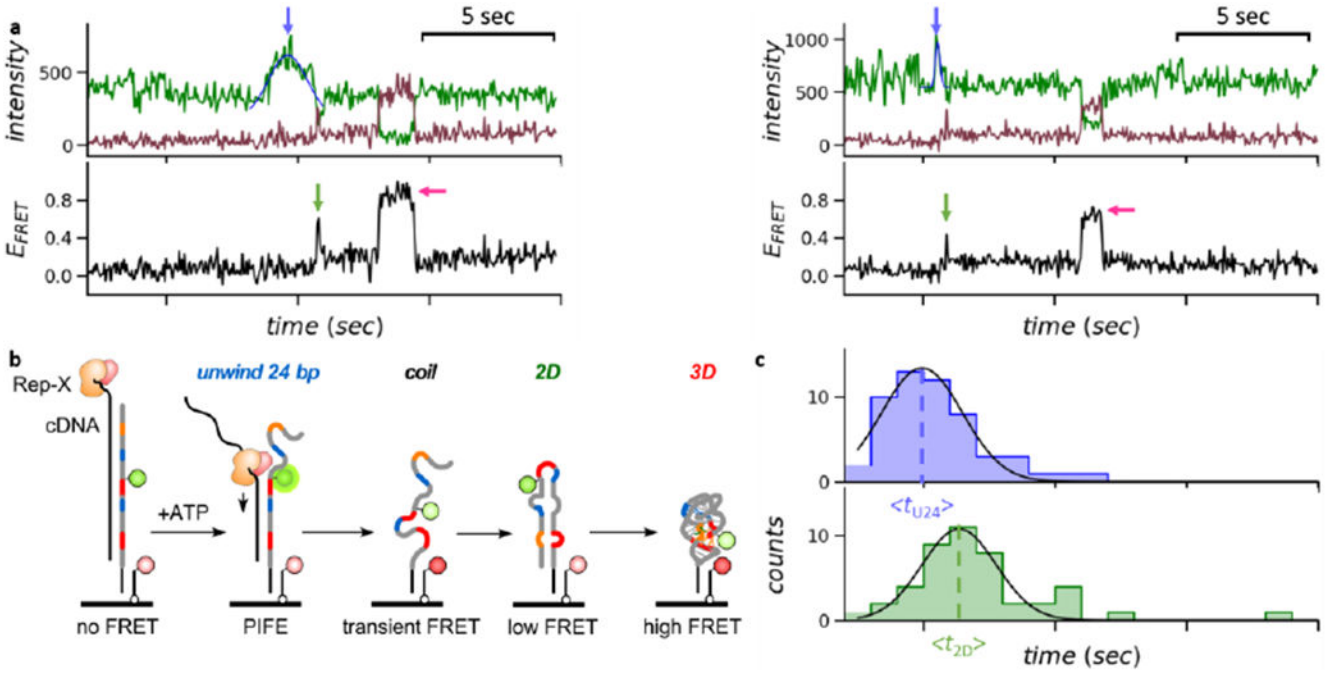


Figure 2.

Real-time observation of the unwinding and folding signals in VFA. (a) FRET trajectories showing the PIFE peaks (blue arrow) and the maiden 2D (green arrow) and 3D (magenta arrow) folding events. The center of each PIFE peak was determined by Gaussian fitting, (b) Schematics of the "order-of-events" in VFA. (c) Time histograms of the PIFE peak centers and the maiden 2D folding transitions, respectively. Gaussian fitting determined the average times (as marked by the dotted lines). The Δt_{2D-U24} was 1.4 s.

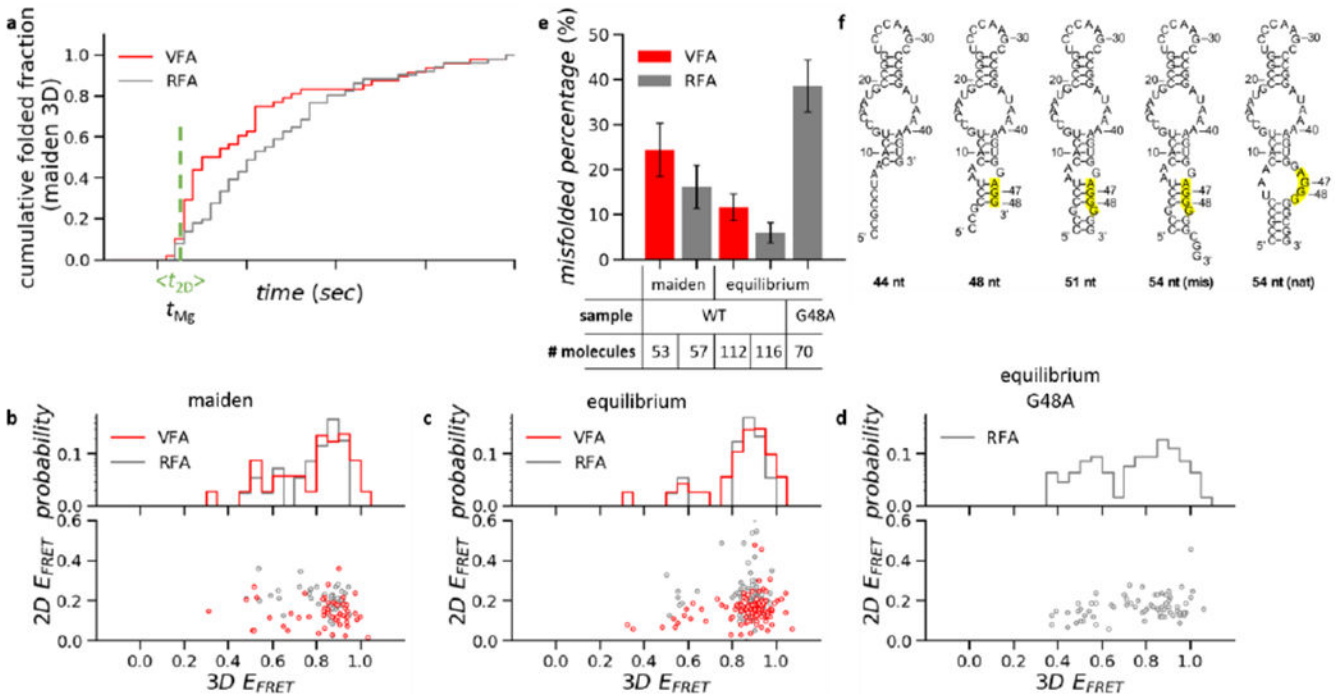


Figure 3.

Characterization of the maiden 3D folding transitions, (a) Cumulative time histograms of all maiden 3D folding transitions (native and non-native). In VFA, molecules that remained as heteroduplex after unwinding buffer addition could not undergo 3D folding transitions (Figure lc middle), therefore they were not included in this analysis. After rejecting the heteroduplex population, the cumulative folded fraction in VFA reaches the same level as RFA (normalized to 1.0). The vertical line indicates the average time of the maiden 2D folding transitions in VFA and the time of $MgCl_2$ addition in RFA. (b–d) Transition scatter plots of E_{FRET} before vs. after 3D folding transitions (bottom) and E_{FRET} histograms of 3D structures (top) from the maiden folding events of the wild type (b), from the equilibrium folding events of the wild type (c) and G48A mutant (d). (e) The percentage of misfolding in various conditions. Error bars represent the standard error estimated by bootstrapping, (f) Simulation of the co-transcriptional 2D folding vs. the number of nucleotides synthesized. The high-lighted region represents the blocked nucleotides that otherwise form PK2.