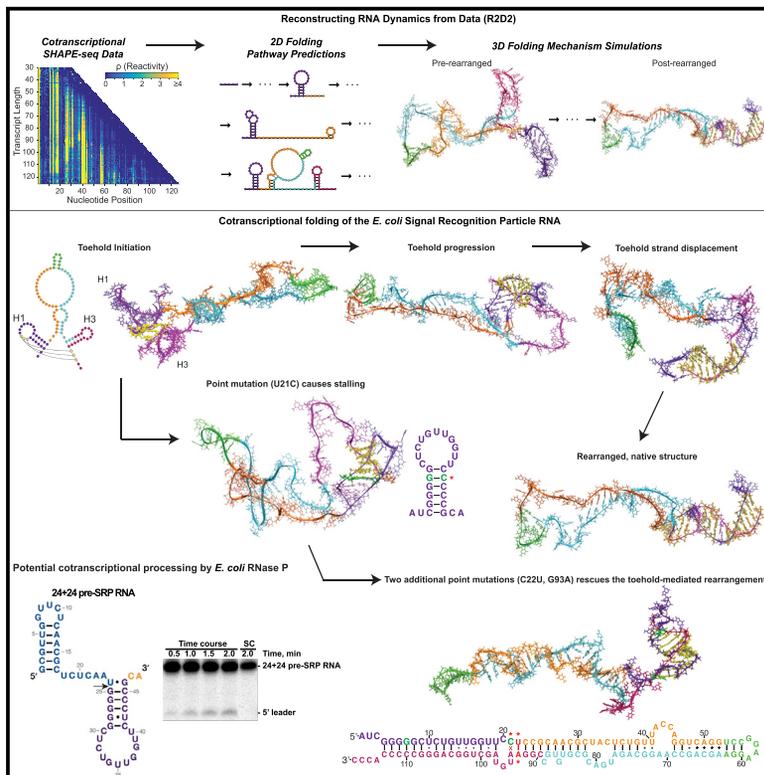


Computationally reconstructing cotranscriptional RNA folding from experimental data reveals rearrangement of non-native folding intermediates

Graphical Abstract



Authors

Angela M Yu, Paul M. Gasper, Luyi Cheng, ..., Venkat Gopalan, Alan A. Chen, Julius B. Lucks

Correspondence

achen6@albany.edu (A.A.C.), jblucks@northwestern.edu (J.B.L.)

In Brief

Experimentally guided multi-scale modeling of the cotranscriptional folding pathway of the *E. coli* signal recognition particle (SRP) RNA reveals that it undergoes an internal toehold-mediated strand-displacement process that requires local structural fluctuations. An intermediate of this pathway is a substrate for RNase P, suggesting concurrent transcription and processing.

Highlights

- Computational modeling of cotranscriptional folding pathways from SHAPE-seq data
- *E. coli* SRP RNA needs local structural fluctuations for toehold strand displacement
- Point mutations can disrupt and rescue cotranscriptional RNA folding pathways
- *E. coli* RNase P cleaves a cotranscriptional intermediate of the *E. coli* SRP RNA

Technology

Computationally reconstructing cotranscriptional RNA folding from experimental data reveals rearrangement of non-native folding intermediates

Angela M Yu,^{1,2} Paul M. Gasper,³ Luyi Cheng,⁴ Lien B. Lai,^{5,6,7} Simi Kaur,^{3,7} Venkat Gopalan,^{5,6} Alan A. Chen,^{3,*} and Julius B. Lucks^{2,8,*}

¹Tri-Institutional Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY 10065, USA

²Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60201, USA

³Department of Chemistry and the RNA Institute, University at Albany, Albany, NY 12222, USA

⁴Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60201, USA

⁵Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

⁶Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: achen6@albany.edu (A.A.C.), jbuckles@northwestern.edu (J.B.L.)

<https://doi.org/10.1016/j.molcel.2020.12.017>

SUMMARY

The series of RNA folding events that occur during transcription can critically influence cellular RNA function. Here, we present reconstructing RNA dynamics from data (R2D2), a method to uncover details of cotranscriptional RNA folding. We model the folding of the *Escherichia coli* signal recognition particle (SRP) RNA and show that it requires specific local structural fluctuations within a key hairpin to engender efficient cotranscriptional conformational rearrangement into the functional structure. All-atom molecular dynamics simulations suggest that this rearrangement proceeds through an internal toehold-mediated strand-displacement mechanism, which can be disrupted with a point mutation that limits local structural fluctuations and rescued with compensating mutations that restore these fluctuations. Moreover, a cotranscriptional folding intermediate could be cleaved *in vitro* by recombinant *E. coli* RNase P, suggesting potential cotranscriptional processing. These results from experiment-guided multi-scale modeling demonstrate that even an RNA with a simple functional structure can undergo complex folding and processing during synthesis.

INTRODUCTION

RNA structures begin to form during transcription. A nascent RNA exiting RNA polymerase (RNAP) transitions through intermediate structures that can ultimately influence the RNA's final fold and function (Kramer and Mills, 1981). Because RNA folding generally occurs faster than transcription (Mustoe et al., 2014), the 5' to 3' directionality of RNA synthesis guides a cotranscriptional "folding pathway" (Pan and Sosnick, 2006). Each time an RNA is transcribed, the ensuing order of folding is critical for essential catalytic RNAs to adopt a functional structure, riboswitches to make regulatory decisions, ribonucleoprotein complexes (e.g., ribosome) to assemble, and RNA processing to take place with efficiency and fidelity (Al-Hashimi and Walter, 2008; Saldi et al., 2018; Serganov and Nudler, 2013). Thus, establishing the principles of cotranscriptional folding is important to understand how each RNA adopts its native structure, with additional payoffs for better appreciating

the dynamic behavior that underpins RNA-based molecular machines and switches.

Despite the widespread biological importance, we still lack a complete understanding of the dynamic, nonequilibrium folding pathways that RNAs undergo during transcription. Pioneering RNA-folding studies showed that the synthesis order of RNA sequence elements is important for establishing functional folds of RNA enzymes (Heilman-Miller and Woodson, 2003; Pan et al., 1999). Enzymatic RNA structure probing was then used to generate models of cotranscriptional folding processes (Wong et al., 2007). Single-molecule force spectroscopy has also been used to track in real time the major folding events of regulatory riboswitches (Frieda and Block, 2012). To complement these approaches with higher-resolution structural information, we previously developed cotranscriptional SHAPE-seq, a chemical structure probing method that captures nucleotide-resolution flexibility data for each length of a nascent RNA in stalled transcription elongation complexes (Watters et al., 2016a). While

these experimental methods are powerful, the resulting data are complex and cannot be directly used to obtain specific RNA structure models.

Computational RNA folding algorithms are important tools for generating models of RNA structure and folding. Some of these algorithms modify minimum free energy (MFE) folding calculations to capture some features of cotranscriptional folding (Proctor and Meyer, 2013) or use stochastic simulations of RNA folding with growing chain length to model cotranscriptional folding (Danilova et al., 2006; Geis et al., 2008; Hofacker et al., 2010; Xayaphoummine et al., 2005). Comparative methods utilizing multiple sequence alignments and evolutionary trees have also been developed to capture potentially conserved transient structures (Wiebe and Meyer, 2010).

We developed a method, named reconstructing RNA dynamics from data (R2D2), that combines nucleotide-resolution experimental RNA structure chemical probing data with computational structure prediction algorithms to reconstruct models of secondary and then tertiary RNA cotranscriptional folding pathways. We applied R2D2 to model the folding pathway of the *Escherichia coli* signal recognition particle (SRP; or 4.5S) RNA, a highly conserved non-coding RNA that is found in all kingdoms of life (Rosenblad et al., 2009). The bacterial SRP RNA binds to the Ffh protein to form the SRP, which recognizes nascent signal peptide sequences and delivers ribosome-nascent chain complexes to the inner membrane for translocation through docking to the SRP receptor. The SRP RNA, which fulfills this critical role in cellular protein biogenesis, consists of a single long hairpin containing several internal bulges and non-canonical base pairs (Batey et al., 2000). This tertiary fold is thought to be generated prior to removal of a 5' leader sequence by RNase P (Bothwell et al., 1976), an essential endonuclease known primarily for its role in tRNA 5'-maturation (Altman, 2007). The *E. coli* SRP RNA is a valuable model for studies of nascent RNA folding, because previous studies indicate that during transcription, the SRP RNA rearranges from an intermediate hairpin fold that differs substantially from the single long hairpin into an extended helical structure (Wong et al., 2007) that resembles the functional structure (Hsu et al., 1984; Jomaa et al., 2017). We therefore applied R2D2 to this model system to uncover mechanistic insights into this rearrangement process.

Our secondary structure models of the processed form of the SRP RNA confirmed the overall rearrangement and inspired a point mutation within an intermediate hairpin that disrupts the cotranscriptional rearrangement of the SRP RNA. We then performed all-atom molecular dynamics (MD) simulations to assess possible mechanisms for the native sequence rearrangement and to gain insights into how a single mutation can disrupt this process. Upon evaluating multiple rearrangement mechanisms, the simulations suggest that the rearrangement likely proceeds via an internal toehold-mediated strand-displacement mechanism. This folding route requires local structural fluctuations within the intermediate hairpin, and the point mutation abolishes these fluctuations. We also engineered point mutations that were predicted to reintroduce flexibility into the intermediate hairpin, and as expected, such a change rescued the ability of the SRP RNA to cotranscriptionally rearrange into its native fold. The presence or absence of the 5' leader that is cleaved by RNase

P was not found to affect these folding mechanisms. Interestingly, our models predicted that one of the intermediates could serve as a natural substrate for RNase P. Indeed, we found that the intermediate is cleaved *in vitro* by recombinant *E. coli* RNase P, suggesting that SRP RNA processing could occur cotranscriptionally as well as the established post-transcriptional pathway. While this work was being performed, several of our structural predictions were corroborated by an independent study that used a high-resolution optical tweezers instrument to follow in real time and on a single-molecule scale the cotranscriptional folding of the same SRP RNA sequences (Fukuda et al., 2020).

Overall, this work presents a method for multi-scale modeling of RNA cotranscriptional folding pathways from experimental data and uncovers efficient ways by which RNAs exploit toehold-mediated strand-displacement mechanisms to rearrange intermediate structure to final folds.

DESIGN

While *in silico* cotranscriptional folding predictors show great promise, the algorithms could benefit from high-resolution experimental studies to corroborate, guide, and improve their predictions. For example, RNA chemical probing data can be used as restraints in computational RNA folding algorithms to improve the agreement between equilibrium predictions and experimental measurements (Deigan et al., 2009). However, these algorithms were developed to model RNA folding under equilibrium conditions, and efforts to predict cotranscriptional secondary structure folding pathways from chemical probing data are in early stages (Li and Aviran, 2018). To address this gap, we developed R2D2 to implement experiment-guided multi-scale modeling of RNA cotranscriptional folding.

R2D2 uses nucleotide-resolution chemical probing data as input to reconstruct models of secondary and then tertiary RNA cotranscriptional folding pathways (STAR methods). Secondary structure modeling for each length of a growing nascent RNA begins by first sampling possible structures using RNA sequence information and cotranscriptional SHAPE-seq data. Subsequently, using an optimized structure-to-data distance metric, sampled structures that are most consistent with the experimental data are selected, resulting in a family of possible structural states reflecting intermediate nascent RNA lengths generated during transcription. This design choice was inspired by previous methods (Ding et al., 2004; Ouyang et al., 2013; Quarrier et al., 2010) that pioneered 2D RNA structure sampling and selection. However, R2D2 differs in the sampling and selection protocols, and importantly connects secondary structures to 3D dynamic modeling. R2D2's secondary structure reconstruction is then used as a starting point for all-atom MD simulations to generate 3D models of cotranscriptional folding transitions observed between specific predicted intermediate states.

Most prior approaches to simulate cotranscriptional RNA folding operate purely at the secondary structure level (Danilova et al., 2006; Geis et al., 2008; Hofacker et al., 2010; Xayaphoummine et al., 2005) and are therefore unable to capture the inherent 3D nature of topological strain, multi-helix junctions, and long-range base pairs, including pseudoknots and kissing loops.

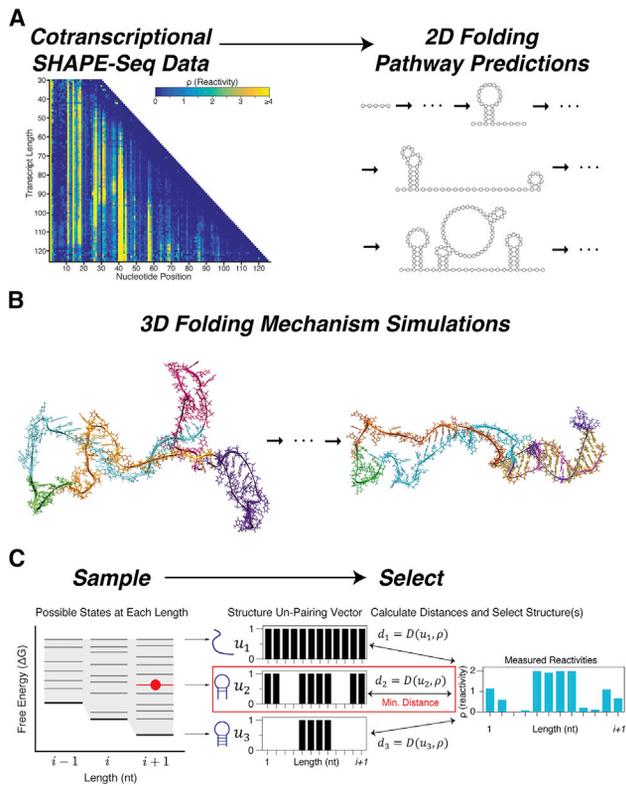


Figure 1. Overview of the reconstructing RNA dynamics from data (R2D2) approach

(A) Schematic of the secondary structure prediction method in R2D2. Cotranscriptional SHAPE-seq is first used to determine reactivities at each nascent transcript length. These reactivities are used to generate secondary (2D) structures along these transcript lengths. Color scalebars represent color-coded ρ SHAPE-seq reactivities for each nucleotide.

(B) 3D simulations are then used to determine the feasibility of structural transitions between specific states within the ensemble of 2D predictions.

(C) Outline of the secondary structure prediction method. Potential RNA structures are statistically sampled for every nascent RNA length. For each length, structures are tested for consistency with the reactivity data at that length, and the most consistent structure is selected. This process is then repeated multiple times for each dataset to obtain a collection of structures over all of the nascent RNA lengths that represent structures along the cotranscriptional folding pathway that are consistent with the data. See also [Table S1](#).

Given the disproportionate impact such interactions have on the kinetics of cotranscriptional folding of SRP, a 3D model was clearly needed for this study. Conversely, 3D simulations have been used previously to study the folding pathways of the SAM-1 riboswitch ([Whitford et al., 2009](#)), utilizing a Gō-like energy model originally developed for studying protein folding. Our approach shares the motivation of reducing folding frustration; however, the Gō-model approach requires a solved 3D structure as an input, which is unavailable for the SRP pre-rearrangement complex. Furthermore, the implicit solvent- and native-contact-based potential cannot be expected to accurately depict folding intermediates stabilized by non-native tertiary interactions. Therefore, there is a need for a new approach that preserves the overall efficiency of the Gō-like models even

while retaining the general applicability of a traditional explicit solvent-based MD simulation. We accomplished this objective by incorporation of selective R2D2-derived restraints applied to all-atom, explicit solvent model simulations capable of the *de novo* folding of small RNAs.

RESULTS

A sample-and-select approach to reconstructing RNA folding pathways from experimental data

We developed a method to merge computational multi-scale RNA structure algorithms with nucleotide-resolution datasets generated from cotranscriptional SHAPE-seq experiments that probe nascent RNA structure ([Figure 1](#)). Cotranscriptional SHAPE-seq begins with *in vitro* transcription of a DNA template library that directs the synthesis of each intermediate length of a target RNA using RNAP roadblocks ([Watters et al., 2016a](#)). Transcription from this template library generates nascent RNAs of all intermediate lengths of the target sequence, which are rapidly probed with the fast-acting SHAPE reagent benzoyl cyanide (BzCN; self-inactivation $t_{1/2}$ of 250 ms) to covalently modify the RNA according to its structure ([Mortimer and Weeks, 2007](#)). RNA nucleotides that are unconstrained by secondary or tertiary structure are more reactive and easily modified ([Aviran et al., 2011](#); [Bindewald et al., 2011](#)). Library preparation, sequencing, and bioinformatics analysis is then used to generate SHAPE reactivities for each nucleotide of each nascent intermediate-length RNA species ([Watters et al., 2016a](#)) ([Figure 1A](#)).

We used a sample-and-select method to reconstruct secondary structure folding intermediates within R2D2. The R2D2 sample-and-select method consists of two steps: (1) generate a set of possible structures at each nascent RNA length by sampling candidate structures from the sequences alone, and (2) computationally select the most likely structure(s) using observed experimental data ([Figure 1C](#)). Thus, R2D2 requires SHAPE-seq data to select structures, and SHAPE-seq data are a necessary input. Comparison between SHAPE-seq reactivities and sampled structures with a “distance” metric that reflects how similar reactivity patterns are to candidate secondary structures is used to select structures that are most consistent with the data at each nascent RNA length ([Figure 1C](#); [Table S1](#)).

To generate candidate structures, the sample method statistically examines structures with a large sample size using the partition and stochastic functions of the RNAstructure suite of computational secondary structure prediction tools ([Reuter and Mathews, 2010](#)). We applied three variations of the partition method that incorporated experimental SHAPE reactivities in different ways to sample 150,000 structures for each length to increase the diversity of structures sampled ([STAR methods](#)).

To select structures from this sampled set, we developed six metrics to calculate the distance between a given SHAPE-seq reactivity spectrum and a given RNA secondary structure ([Table S1](#); [STAR methods](#)). Structures with the minimum distance calculated from the reactivity spectrum were selected from a candidate sampled set ([Figure 1C](#)). By applying this selection at each nascent RNA length, we could reconstruct possible folding intermediates that were most consistent with the experimental data.

Benchmarking sample-and-select on equilibrium refolding data

We next assessed the accuracy of each proposed distance metric. As there are currently no established benchmarks for cotranscriptional folding predictions, we instead assessed distance metrics by predicting the equilibrium folds of an established benchmark panel of RNAs using SHAPE-seq data (Loughrey et al., 2014). Each distance metric contains several parameter values that are used to determine how the SHAPE reactivities are compared to sampled RNA structures; ρ_{max} and ρ_c determine cutoffs in reactivity values, and α weighs the contributions from paired versus unpaired positions (STAR methods). For each of the six distance functions, we determined the optimal values of the three fit parameters by applying the sample-and-select method to a panel of RNAs previously used to benchmark equilibrium SHAPE-directed secondary structure prediction algorithms (Table S1). The best-performing parameter sets performed comparably to the SHAPE-data-based output of the fold module of RNAstructure, a widely used RNA secondary structure prediction algorithm (Table S2; STAR methods).

Reconstructing the secondary structure cotranscriptional folding pathway of the *E. coli* SRP RNA sequence

We next applied the R2D2 sample-and-select method to our previously published SRP RNA cotranscriptional and equilibrium-refolded SHAPE-seq datasets (Watters et al., 2016a), and this study further characterizes mutants designed based on R2D2 results. The SRP RNA sequence used was based on Wong et al. (2007), who examined SRP RNA folding using a variant that has AUC in place of the 5' native 24-nt leader. Before applying R2D2's sample-and-select to this cotranscriptional probing dataset, we removed the last 14 nt from the 3' end of each RNA sequence to account for the RNAP footprint (Komissarova and Kashlev, 1998). To compare cotranscriptional to equilibrium-refolded datasets that do not contain an RNAP footprint, we compared trimmed cotranscriptional transcript lengths to equal lengths of the equilibrium-refolded RNA sequence from each experimental dataset. To visualize R2D2 predictions at each nascent RNA length, we plotted the computed free energies (ΔG) of each selected structure and connected all possible paths between selected structures for visual convenience, noting that connections do not imply transition probabilities between states (Figure 2A). Notably, we observed that distinct structures can have the same minimum distance to the experimental data, which may reflect a mixed population of RNA states at specific lengths. We therefore chose to leave these multiple structures as distinct possibilities that are equally consistent with the data.

Despite diversifying our sampling procedure 150-fold over some previous sample-and-select methods (Ding et al., 2004; Ouyang et al., 2013), we found that it is intractable to generate an exhaustively complete set of candidate structures at each length due to the slow convergence of the stochastic sampling method (Figures S1A and S1C). Thus, iterations of sample-and-select may generate different sets of candidate structures that can be consistent with the data. To incorporate this variability in sampling, we ran 100 iterations of R2D2 sample-and-select on

each SHAPE-seq dataset to generate a family of possible intermediate folding states (Figure 2A). We applied this method to cotranscriptional SHAPE-seq datasets of the SRP RNA sequence, as well as SHAPE-seq datasets from experiments performed on an equilibrium-refolded population of the same SRP RNA sequence intermediates to compare out-of-equilibrium to equilibrium predictions of intermediate states (Figures 2 and S2). Overall, we observed that cotranscriptional and equilibrium predictions are similar for short RNA lengths, diverge as the RNA length increases, and ultimately converge near full length.

To analyze structural changes that may occur during transcription, we extracted specific structures chosen by the select method at each nascent length. We viewed the family of selected structures at each length using RNAbow software (Aalberts and Jannen, 2013), which revealed specific structural changes across the SRP RNA folding trajectory that differ between out-of-equilibrium and equilibrium datasets (Figures 2B–2H, S2B–S2E, and S2G–S2J). When the first 23–25 nt are free to fold in the cotranscriptional SHAPE-seq predictions, we detect the formation of a 5' helix containing ≥ 3 bp that persists through most of the folding pathway (Figures 2B–2F). Interestingly, this 5' helix differs in its makeup from the 5' helix consisting of positions 4–10 paired to 16–22 that was inferred based on enzymatic probing experiments (Figures S3A and S3C) (Wong et al., 2007). Instead, we consistently predict a 5' helix 1 (H1) where positions 3–8 are paired to 20–25, which is consistent with the previous enzymatic probing results (Figures S3B and S3D) and our cotranscriptional SHAPE-seq data (Figure S3E). We found that H1 is present for a large portion of the folding pathway and, based on our reconstructed states, starts to rearrange into the native long helical structure at lengths 110–111 nt (Figures 2B–2G, S2B–S2D, and S2G–S2I; Video S1).

The next highly persistent structure that forms is a helix created when nucleotides 53–55 pair with 60–62 to form the apical stem-loop of the native structure (Figures 2C–2G, S2B–S2E, and S2G–S2J). We note, however, that our reconstructions do not predict the formation of four non-canonical interactions that are present in the crystal structure of the *E. coli* SRP RNA: C49–A66, A50–C65, G51–G64, and G52–A63 (Batey et al., 2000). We attribute this limitation to the reliance of R2D2's sample-and-select method on RNAstructure's partition and stochastic functions, which are not able to sample structures that contain non-canonical interactions, although portions of the cotranscriptional SHAPE-seq reactivity matrix in this region show elevated reactivities, indicating this region also likely does not close on the 30-s timescale of our experiment. Despite these differences, R2D2 does reconstruct most of the mature SRP RNA sequence structure by length 117 (Figures 2H, S2E, and S2J).

Prior to folding into the final structure, the sample-and-select method also predicts 3' hairpin structures at various transcript lengths. One such structure is between nucleotides 72 to 90, which we denote early helix 3 (eH3), and the next is between nucleotides 87 to 105, which we denote helix 3 (H3) (Figures 2E and 2F). Both eH3 and H3 locally insulate bases that form different pairs with nucleotides that are ultimately sequestered within H1 in the final structure. H3 was previously found by comparative analysis of SRP RNA sequences from diverse bacterial species, suggesting it may be an evolutionarily conserved transient

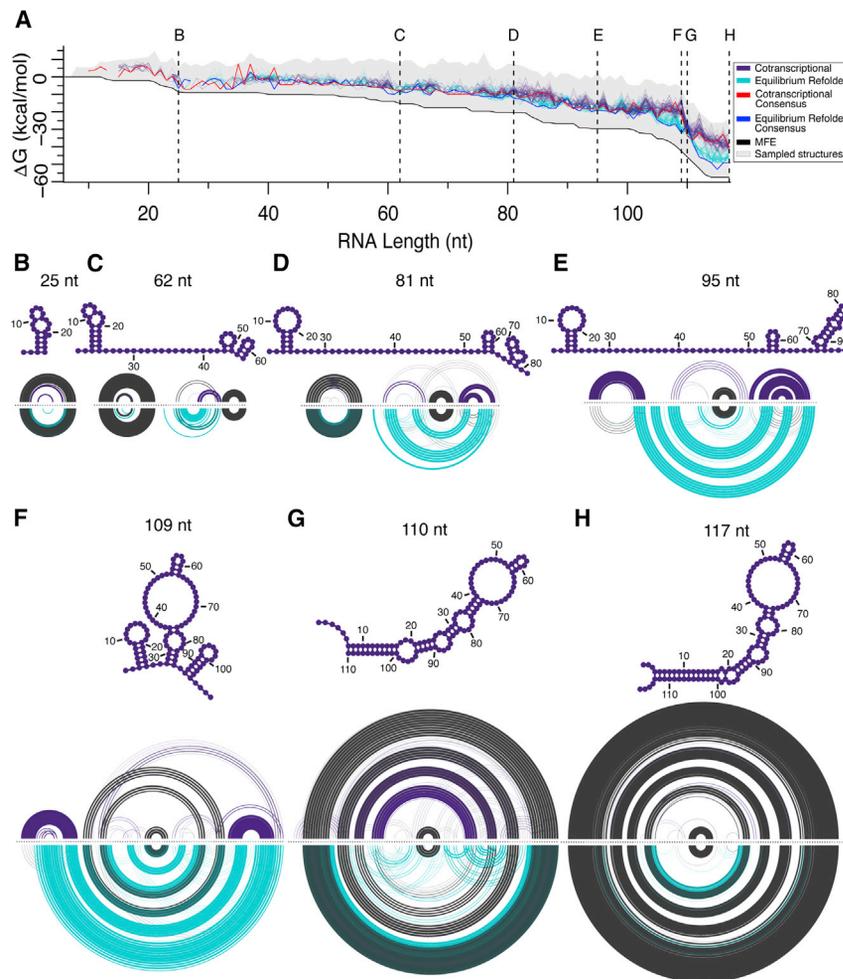


Figure 2. R2D2 2D pathway predictions for the *E. coli* SRP RNA sequence

(A) Secondary structure predictions by R2D2 on cotranscriptional and equilibrium-refolded SHAPE-seq data of the *E. coli* SRP RNA sequence. For each dataset, 100 folding pathway predictions were performed and plotted according to the free energy (ΔG) of the RNA structures predicted along the cotranscriptional (purple) or equilibrium-refolded (turquoise) pathway. The range of ΔG values sampled is represented by gray shading, while the ΔG of chosen structures are represented by dots. For visual convenience, dots are connected by lines to view possible free-energy changes along the folding trajectory. Consensus structure lines connect ΔG of structures containing base pairs that occur in over 50% of the 100 iterations performed on the cotranscriptional (red) and equilibrium-refolded (blue) SHAPE-seq data. Black line connects the minimum free energy (MFE) structures in the sampled set.

(B–H) Seven lengths of 2D predictions by R2D2 are highlighted: (B) 25 nt, (C) 62 nt, (D) 81 nt, (E) 95 nt, (F) 109 nt, (G) 110 nt, and (H) 117 nt. One hundred selected structures are represented as RNAbow plots with base pairs drawn as arcs and the arc thickness indicating prevalence of the base pair among the selected structures. Colored arcs show base pairs that are more frequent in either cotranscriptional (purple) or equilibrium (turquoise) predictions, while gray arcs show base pairs that are shared. The consensus structures from cotranscriptional predictions are shown above each RNAbow plot. We compare equal lengths of the RNA sequence that is free to fold from each experimental dataset. Data plotted in this figure are from cotranscriptional SHAPE-seq replicate 1. See also [Figures S2](#) and [S3](#) and [Video S1](#).

structural feature of the SRP RNA (Zhu et al., 2013). The presence of H1 and H3 present a significant structural barrier to cotranscriptional folding in that both must be broken to form the mature extended helical fold. We note, however, that H3 and eH3 are not predicted in every selected structure, indicating that additional folding pathways are likely.

Sample-and-select models differ from approaches that do not use experimental data

Based on the ΔG folding trajectory, R2D2's sample-and-select chooses structures that are higher in free energy than the MFE predicted with or without experimental data at almost all lengths (Figures 2A, S2A, S2F, and S4A). Other than MFE approaches, one of the most widely used is KineFold, which simulates cotranscriptional folding given only an input sequence and a desired transcription rate (Xayaphoummine et al., 2005). In a comparison between 100 repetitions of KineFold and R2D2, KineFold predicted different folding pathways. The key differences between the two approaches pertain to predictions of transient helices such as H1 and H3 as well as the location of structural rearrangements (Figures S4B–S4F). Predictions between R2D2 and KineFold differ even when simulating 40 s of transcription and stalling

at each intermediate length with KineFold to test if the RNAP roadblocking strategy in cotranscriptional SHAPE-seq (30 s of transcription followed by SHAPE probing) explains differences between R2D2 and KineFold (Figures S4G–S4J).

A single point mutation delays the cotranscriptional rearrangement of the *E. coli* SRP RNA sequence

R2D2 predictions show structural variation within H1 across the folding pathway (Figures 2, S2, and 3A), which we hypothesized is due to the GU pair within the otherwise GC-rich H1. We therefore mutated the native U21 to C21 to change the GU to a GC bp, thereby increasing the stability of H1 and disfavoring the rearrangement into the final helix structure (Figure 3B). R2D2 analysis of the cotranscriptional SRP RNA U21C dataset predicts the presence of H1 at all lengths of the folding pathway through lengths 112, 111, and 110 nt in the first, second, and third replicate, respectively (Figures 3D–3H). In contrast, R2D2 shows the rearrangement into the final extended structure in U21C equilibrium-refolded data occurring earlier at length 109 nt. These differences in R2D2's 2D results of SRP RNA U21C cotranscriptional and U21C equilibrium-refolded SHAPE-seq reactivities are due to reactivity differences (e.g., consistent drops in H1 loop

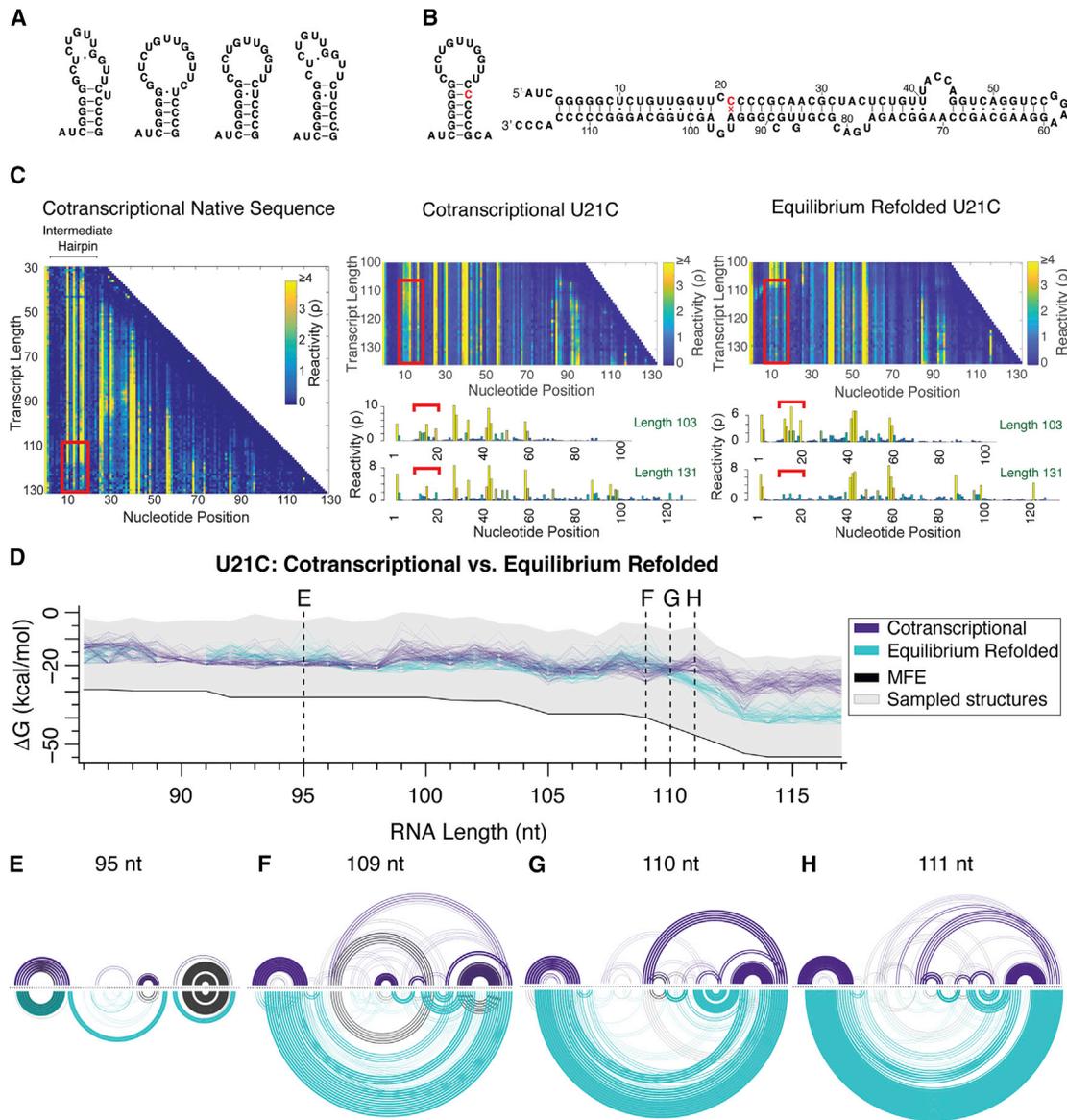


Figure 3. A single point mutation disrupts cotranscriptional rearrangement of the mature *E. coli* SRP RNA

(A) Examples of H1 variability in 2D predictions in the folding pathway of the mature sequence indicate potential flexibility.

(B) Schematic of the SRP RNA U21C mutation in H1 and the full-length secondary structure.

(C) Cotranscriptional SHAPE-seq reactivities from the mature (left) sequence show drops in reactivities (red box) toward the end of the folding pathway. Color scalebars represent color-coded ρ SHAPE-seq reactivities for each nucleotide. The reactivity matrix for replicate 1 of the SRP RNA U21C sequence (middle) has generally higher reactivities in these positions throughout, while equilibrium-refolded SRP RNA U21C SHAPE-seq data (right) contains decreases in reactivities in this region. Plotted below these matrices are their respective reactivities from transcript lengths 103 and 131 with H1 loop reactivities under the red bracket.

(D) Trajectory plot of R2D2 predictions for the U21C sequence following Figure 2.

(E–H) Structures from four lengths are highlighted in RNAbow plots: (E) 95 nt, (F) 109 nt, (G) 110 nt, and (H) 111 nt.

See also Figure S5.

reactivities at lengths 108–109 nt of the equilibrium-refolded data; Figure 3C). Detection of unknown events with tunable thresholds (DUETT), a recently developed algorithm to systematically detect reactivity changes in cotranscriptional SHAPE-seq datasets (Xue et al., 2019), detected these drops in reactivity.

The lack of H1 predictions at near full-length RNAs (Figures S5A and S5B) indicate that rearrangement of H1 is possible given the

experimental data but is delayed due to minimization of local fluctuations in H1 (discussed below in [Uncovering potential mechanisms of the SRP RNA cotranscriptional structural rearrangement with all-atom simulations](#)). However, we also explore the possible limitations in the Boltzmann-distribution-directed sampling methods used as a reason for the predicted cotranscriptional rearrangement of U21C. Boltzmann-distribution-directed sampling

is naturally biased toward sampling lower free energy structures, making it difficult for the algorithm to choose out-of-equilibrium structures, especially those with increasing RNA lengths. To investigate this possibility, we added to the selection pool structures sampled from the previous six lengths and extended them with unpaired nucleotides. With these additional structures, we find that H1 persists through lengths 113–116 nt, while rearrangement is predicted at length 117 nt in only two of the three replicates (Figures S5A and S5B). We also ran R2D2 using this modified sampling procedure on the native SRP RNA sequence as a control and found lengths 115–117 nt (Figures S5A and S5B) are predominantly predicted to be rearranged as expected. Application of the standard R2D2 sample-and-select procedure to the SRP U21C equilibrium-refolded datasets showed the presence of H1 but recovered the rearrangement into the final extended helical structure after length 109 nt (Figures 3D–3H). Taken together, these data demonstrate that a single point mutation can delay a key transition of the SRP RNA cotranscriptional folding pathway and kinetically trap the RNA in non-native intermediate structures.

A single GU wobble is critical for the *E. coli* SRP RNA cotranscriptional rearrangement into the extended final fold

Since the replacement of a single GU bp in the predicted H1 helix is enough to disrupt the cotranscriptional rearrangement of SRP RNA, we sought to test the idea that reintroduction of a GU pair in H1 would rescue the cotranscriptional rearrangement. We therefore designed a mutation (U21C, C22U, and G93A) that reintroduces a GU wobble pair one position lower in the stem of H1 and maintains sequence complementarity between nucleotides 22 and 93 (Figure 4A). The cotranscriptional SHAPE-seq reactivity matrix for this mutant shows a drop in reactivities at length 119 nt (~105 nt free to fold), which was determined with automated detection of reactivity changes (Xue et al., 2019) (Figure 4B). In addition, when applied to this dataset, R2D2 sample-and-select predicts that this mutant follows a similar folding pathway as the native sequence (Figures 4C and 4D) and that the rearrangement occurs at lengths 110 and 111 nt between three replicates, which is the same rearrangement length observed in the wild-type sequence (Figures 4E and S5F). Overall, these data point to the critical requirement of a GU pair within H1 to facilitate the cotranscriptional rearrangement into the final extended helix structure.

Uncovering potential mechanisms of the SRP RNA cotranscriptional structural rearrangement with all-atom simulations

We next sought to determine the mechanism by which the SRP RNA rearranges during transcription, and the role of the H1 GU bp in this process. Paradoxically, H3 would be expected to impede this rearrangement, as both H3 and H1 need to somehow unzip and hybridize together to form the native extended helix structure. We therefore focused on mechanisms by which the three-hairpin consensus structure at 109 nt of cotranscriptional SHAPE-seq replicate 1 (Figure 2F) can rearrange into the extended helix structure at 110 nt (Figure 2G). Four distinct potential transition pathways were identified: the inside-out (Figure 5A), kissing-loop (Figure 5B), late-toehold (Figure 5C), and

early-toehold (Figure 5D) pathways. We used all-atom MD simulations to characterize the relative feasibility of each of the four proposed transition pathways from the stable folding intermediate containing H1 and H3 (Figure 6A) to the mature fold (Figure 6B). Each pathway suggests that the rearrangement mechanism begins with a different set of base-pairing interactions (Figure 5). To test each pathway, weak attractive biasing forces between specific nucleotides were sequentially added in a specific order, starting at the initial proposed interaction to facilitate transitioning to the mature fold. Eight replicate simulations were performed for each path (STAR methods).

The inside-out hypothesis involves simultaneously breaking H1 and H3 at their stems from the middle radiating outward to initiate the formation of the native helix (Figure 5A; Video S2). While it was technically possible to observe the inside-out pathway in the simulation with large biasing forces, it would be extremely thermodynamically unfavorable, since it would involve breaking 2 bp for every 1 bp formed for a significant portion of the pathway (Figure 5E; Video S2). This pathway was possible only when stronger restraints were used, thus identifying an upper limit for the strength of the restraints for all of the other transition pathways.

The kissing-loop pathway assumes that bases 17–19 of the H1 loop and bases 98–100 of the H3 loop form an initial base pair to seed the rearrangement (Figure 5B). These nucleotides were chosen because the resulting CG/GU/AU base pairs would produce a significantly stronger kissing complex than those composed of only a GU/AU base pair, analogous to the 2-bp kissing complex that drives Moloney murine leukemia virus (MMLV) genome dimerization (Zhu et al., 2013). The kissing loop was not able to form in all simulations of this pathway, even when each simulation was extended multiple times for 100 ns and the strength of the long-range restraints were doubled (Figure 5E). The mismatch in length of the two helical segments effectively prevents the bases from forming hydrogen bonds in the pre-transition secondary structure.

Finally, the late-toehold pathway assumes that bases 106–108, predicted to be in an unpaired strand at the 3' tail of the RNA at the base of H3, initially pair with bases 9–11 in the H1 loop and form a “toehold” interaction (Figure 5C; Video S3). The initial toehold contacts were found to reliably form in six out of eight attempts as the 3' tail of the nascent RNA is flexible and long enough to reach the loop of H1 (Figure 5E). All simulations that formed the initial toehold contacts proceeded through the refolding pathway to the 110-nt structure.

A decisive advantage of the toehold mechanism is the favorability of the strand exchange process that proceeds in a break-one-form-one-base-pair manner. Once identified as a plausible mechanism, we realized that this toehold-mediated strand displacement can also be initiated earlier in the folding trajectory before H3 forms (Figure 5D). Simulations of the “early toehold” indicate that the absence of H3 actually speeds up the rearrangement due to the greater flexibility of the longer single-stranded 3' tail, the lack of an energetic barrier posed by H3 (Figure 5D), and the increased number of bases available to form the initial toehold. Thus, the toehold-mediated strand-displacement mechanisms are much more plausible than the other pathways considered.

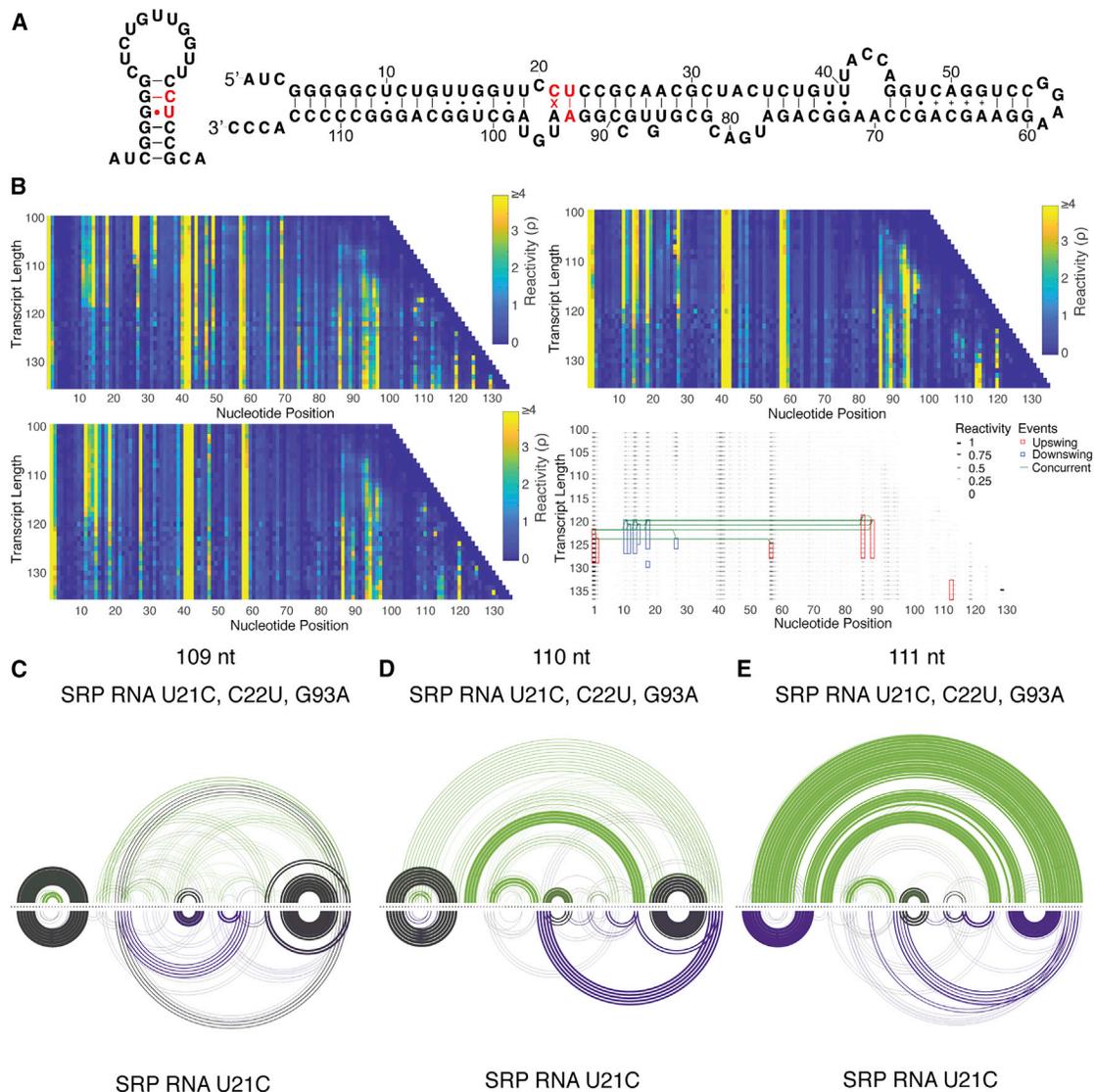


Figure 4. Rescue mutant of SRP RNA U21C confirms the importance of flexibility in H1

(A) Schematic of the rescue mutant U21C, C22U, and G93A overlaid on H1 and the native full-length structure. The rescue mutant introduces a GU base pair in the SRP RNA U21C H1 structure.

(B) Cotranscriptional SHAPE-seq reactivities from SRP RNA U21C, C22U, G93A: replicate 1 (left, top), replicate 2 (left, bottom), and replicate 3 (right, top). Color scalebars represent color-coded ρ SHAPE-seq reactivities for each nucleotide. DUETT analysis (right, bottom) detected downswings (blue) and upswings (red) in reactivity. Events occurring up to two transcript lengths apart are indicated with green lines.

(C–E) RNAAbow plots of SRP RNA U21C, C22U, G93A replicate 1 (green and top) and U21C replicate 1 (purple and bottom) R2D2 predictions following Figure 2. Three lengths are highlighted: (C) 109 nt, (D) 110 nt, and (E) 111 nt.

See also Figure S5.

A detailed examination of the productive toehold-mediated folding pathways reveals several key architectural features that facilitate the rapid folding transition (Video S3). Extension of the initial toehold-seeding interaction to the full rearrangement requires fluctuations from the 11-nt loop of H1 into the stem (Figure 6C). H3, which is weaker than the GC-rich H1, readily unfolds in the simulations after the first few base pairs are formed, and the resulting increase in single-strandedness further facilitates flexibility in hybridization with the H1 loop (Figure 6D). In addition,

the formation of C7-G110 and C8-G109 base pairs requires unraveling of the top of H1's stem, which is facilitated by fluctuations of the GU base pair. Only after C7-G110 and C8-G109 are formed is the H1 hairpin weak enough to open up, allowing the remaining base pairs of the native helix to align and zip up in an energetically downhill process to form the fully extended fold (Figure 6B).

The results described above suggested that the SRP RNA U21C mutant minimizes the ability of H1 to fluctuate, disabling

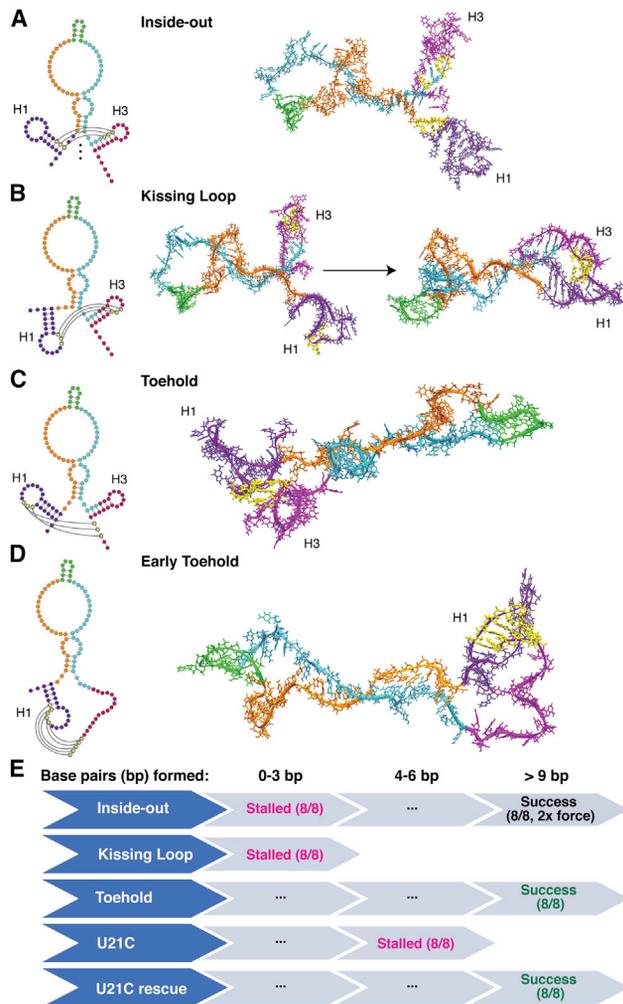


Figure 5. Snapshots of possible rearrangement mechanisms tested by 3D all-atom simulations

R2D2-modeled secondary structures (left) were used as starting points for all-atom MD simulations (right). For each rearrangement that we tested, the pairing interactions that could seed the rearrangement into the native extended hairpin are indicated in yellow. Other nucleotides are colored for visualization, including nucleotides 1–25 (dark purple), 26–52 (orange), 53–62 (green), 63–96 (turquoise), and 97–117 (magenta).

(A) The inside-out hypothesis where H1 and H3 progressively open and convert into the extended hairpin.

(B) The kissing-loop hypothesis where H1 and H3 loops begin the rearrangement process.

(C) The late-toehold hypothesis where nucleotides 106–108 downstream of H3 seed the rearrangement through a toehold with nucleotides 9–11 of H1 loop.

(D) The early-toehold hypothesis where nucleotides 106–110 seed the rearrangement via a toehold with nucleotides 7–11 of H1 loop. The early- and late-toehold hypotheses differ in the structural state of the growing SRP RNA's 3' end before the rearrangement, with the late-toehold hypothesis considering the unraveling of H3.

(E) 3D all-atom simulation trajectory results. Simulations were used to test the potential rearrangement mechanisms in wild-type SRP RNA and test the U21C mutant and its rescue with the late-toehold mechanism. Eight simulations were run for each scenario. Simulations can stall when 0–3 or 4–6 bp form. Otherwise, rearrangement could progress to >9 bp or when twice the force was applied.

See also [Videos S2, S3, and S4](#).

this mutant to efficiently rearrange during transcription. To directly test this hypothesis, we performed simulations of the SRP RNA U21C mutant and found that the toehold can still form between bases 7–110 and 8–109 when restraints were applied, but the mutant cannot transition into the final folded state because of the increased stability of H1 (Figures 6E and 5E). The folding transition still stalled even when double-strength restraints were applied, as these were insufficient to disrupt the G7-C21 base pair. Finally, simulations of the rescue mutant (U21C, C22U, and G93A) confirm that restoring flexibility in the upper stem of H1 recovers the ability to transition to the mature fold, albeit at a slower rate due to the extra base pair that needs to be disrupted to unfold H1 (Figure 6F and 5E; [Video S4](#)).

Overall, our simulations strongly suggest a toehold-mediated strand-displacement rearrangement mechanism that is facilitated by base-pair fluctuations within the stem of H1.

Addition of the precursor sequence to the SRP RNA does not impact rearrangement

We next investigated how cotranscriptional RNA folding could affect the precursor SRP RNA and its processing by RNase P. The precursor SRP (pre-SRP) RNA contains a 5' 24-nt leader sequence, which is thought to contain a small hairpin ([Figure 7A](#)) ([Peck-Miller and Altman, 1991](#)). Interestingly, this precursor hairpin is predicted to fold independently when appended to the shorter SRP RNA lengths that fold into H1 and together form a potential RNase P substrate. We therefore tested if such partial pre-SRP RNA sequences can be processed by RNase P.

For use as a representative substrate, we generated by *in vitro* transcription a pre-SRP RNA (termed 24+24) with the 24-nt leader and the first 24 nt of the mature SRP RNA. Compared to the full-length pre-SRP RNA and pre-tRNAs, the 24+24 pre-SRP RNA substrate differs in two aspects ([Figure 7A](#)). First, it has only 5 bp in the H1 stem, which is shorter than the typical 7-bp acceptor stem of pre-tRNA, and much shorter than the long stem in the full-length pre-SRP RNA. Second, this short substrate has a 2-nt 3'-CA terminus compared to the 3'-CCC of the full-length pre-SRP RNA and the 3'-CCA of pre-tRNAs. Despite these differences, the 24+24 pre-SRP RNA was efficiently cleaved by *in vitro* reconstituted *E. coli* RNase P ([Figures 7A and 7B](#)). Since a K_m of 0.2 μM was calculated for the processing of full-length pre-SRP RNA by *E. coli* RNase P ([Peck-Miller and Altman, 1991](#)), we tested the rate of cleavage of the 24+24 pre-SRP RNA at 2.5 μM to ensure saturation. Our results yielded a turnover number of 5.4 min^{-1} for the 24+24 pre-SRP RNA ([Figure S5G](#)) compared to 37 min^{-1} reported for the full-length counterpart ([Peck-Miller and Altman, 1991](#)).

That the 24-nt leader sequence of the *E. coli* pre-SRP RNA could be cleaved both post- and co-transcriptionally motivated us to examine the effect of the leader sequence on the toehold-mediated strand-displacement mechanism. Additional all-atom simulations were conducted with the 24-nt leader sequence and its hairpin added, which was found not to interfere with the toehold-mediated rearrangement exhibited by the leaderless sequence ([Figures 7C and 7D](#)). In one replicate of the pre-SRP RNA simulations, we observed the order of base pairs displacing the H1 stem to be U21-A94, G25-C90, C24-G91, C23-

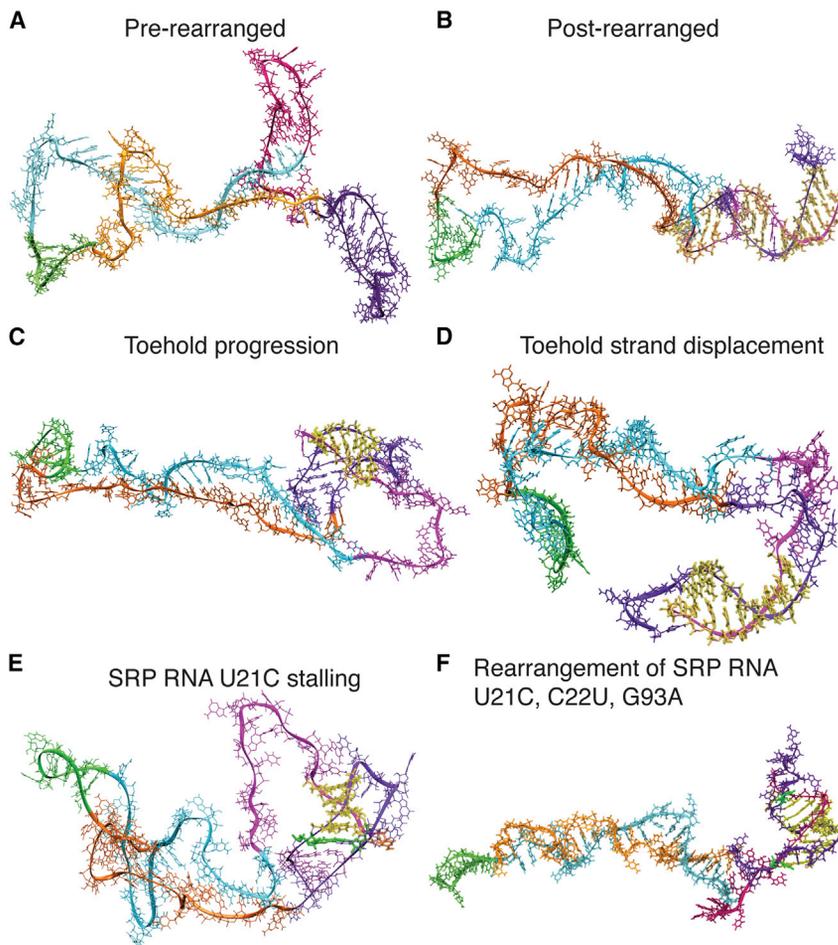


Figure 6. Snapshots of the toehold-mediated rearrangement pathway from MD simulations

(A) Pre-rearranged structure with H1 (purple) and H3 (magenta) present.

(B) Rearranged structure with native base pairs (yellow) forming the extended helix.

(C) Toehold progression to 6 bp of the native helix (yellow) requires unfolding of H3.

(D) Further elongation to a 9-bp native helix (yellow) requires unfolding of H1.

(E) In the SRP RNA U21C mutant, H1 is stabilized by a GC base pair (green) that replaced the GU base pair. Even if a toehold is made to form (yellow), folding stalls, as the G7-C21 base pair cannot be disrupted even when modest biasing forces are applied.

(F) In the SRP RNA U21C, C22U, G93A mutant, the rearrangement can occur and the GC base pair (green) can break.

See also [Video S3](#).

G92, and C22-G93. This finding alludes to the possibility of an ensemble of cotranscriptional folding pathways. Additionally, the U21C mutation (now U42C in pre-SRP RNA) still abrogated toehold-mediated strand displacement and rescued with the addition of C22U and G93A (now C43U and G114A) mutations in simulations.

DISCUSSION

We developed R2D2 to reconstruct nascent RNA folding at high resolution. Our R2D2 analysis of the SRP RNA reveals that although excursions to non-native structures could entail kinetic traps, they may also present a low free-energy path to the final native fold by minimizing structural fluctuations, as revealed here by R2D2 analysis of the SRP RNA. While ribosomal proteins have been shown to modulate rRNA dynamics and therefore the conformational ensemble (Kim et al., 2014), it appears that the same physical principles might help naked cellular RNAs traverse through non-native structures during transcription.

R2D2's secondary structure approach builds on elements in previous RNA folding algorithms but is distinctive in its use of multi-scale modeling to reconstruct out-of-equilibrium folded states along a cotranscriptional folding pathway. Thus, R2D2 is

different from MFE prediction methods, which would not uncover the importance of H1 flexibility because of H1's stability in the SHAPE-directed MFE folding pathway (Figure S3; Video S1). Specifically, the timescale of cotranscriptional folding invalidates the frequently used assumption of equilibrium RNA structural states at each nascent RNA length, making R2D2's combination of experimental data and sample-and-select a promising approach. In this regard, the secondary structure aspect of R2D2 is similar to the

recent SLEQ (Li and Aviran, 2018) and Rsample (Spasic et al., 2018) methods, although the latter are able to additionally estimate population levels of certain RNA structures. Overall, R2D2's merging of multi-scale modeling with experimental data distinguishes it from previous computational methods to study cotranscriptional RNA folding.

We focused our studies on a particular three-helix-containing intermediate structure in the SRP RNA cotranscriptional folding pathway. Using R2D2, we propose that this three-helix structure can efficiently rearrange into a single extended helix through a toehold-mediated strand-displacement mechanism, although we recognize that alternative folding pathways are possible due to the stochastic nature of RNA folding. Even within toehold-mediated mechanisms, multiple toehold-initiation points and rearrangements starting from eH3 or other 3' structures are possible, suggesting various routes to attain the native fold even while centered around a key decision point. The large size of the H1 loop could be important for the increased flexibility of these bases for toehold nucleation as well as exposing a large sequence to capture the many alternate transient 3' end structures.

Overall, it could be that the SRP RNA has evolved to allow multiple toehold-mediated strand-displacement mechanisms to prevent the kinetic folding trap imposed by H1 and H3,

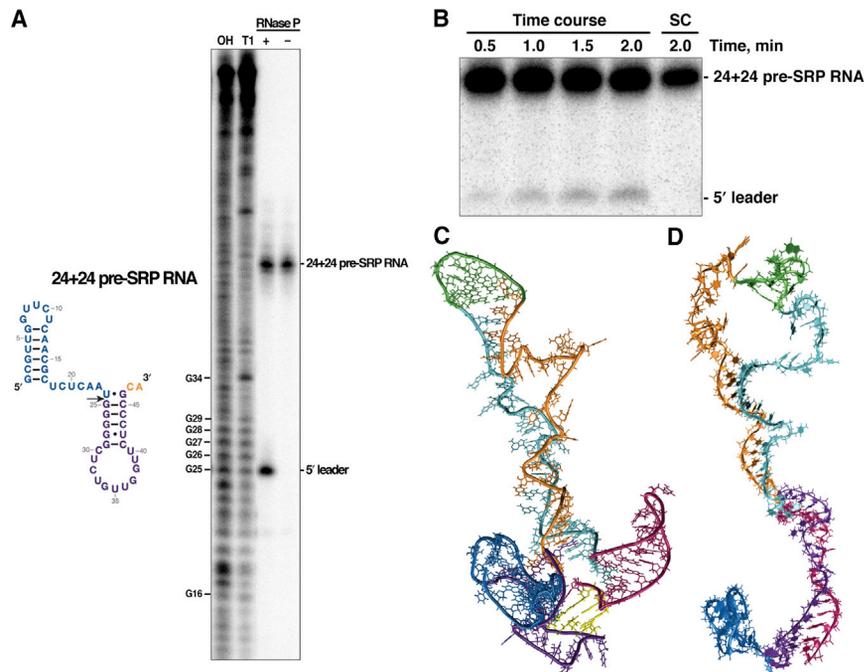


Figure 7. Processing and folding of the pre-SRP RNA

(A) *E. coli* RNase P cleaves the 24+24 pre-SRP RNA correctly at the expected site (arrow). The OH and T1 ladders were generated by alkaline hydrolysis and RNase T1 cleavage, respectively, of the 24+70 pre-SRP RNA. The cleaved 5' leader (blue) migrated with length G25, instead of length U24, because it has a 3'-OH compared to the 2',3'-cyclic phosphate in the RNase T1 products.

(B) A representative gel of the time-course assay used to determine the rate of cleavage of the 24+24 pre-SRP RNA by *E. coli* RNase P. The initial velocities were calculated from three replicates, and the maximal cleavage did not exceed 5% of the total substrate. A turnover number of $5.4 \pm 0.5 \text{ min}^{-1}$ was obtained from these measurements. SC, substrate control incubated without *E. coli* RNase P.

(C) The hairpin in the 5' leader (blue) does not impede the toehold initiation (yellow) in R2D2 3D simulations.

(D) The hairpin in the 5' leader also does not affect the toehold-mediated rearrangement.

See also [Figure S5](#) and [Videos S5, S6, and S7](#).

which were previously identified as potential transient helices that are evolutionarily conserved (Zhu et al., 2013). Recently, it has been shown that many natural RNAs contain long-range interactions in the cell, some of which occur over 1 kb away (Lu et al., 2016). Given the propensity of RNAs to form local structures cotranscriptionally, toehold-mediated strand displacement could be one of the most efficient ways for RNAs to undergo large-scale rearrangements. Detailed *in vitro* studies of toehold strand-displacement reactions have demonstrated rates on the order of $10^6/\text{M/s}$ for a bimolecular strand-exchange reaction (Šulc et al., 2015; Zhang and Winfree, 2009). In addition, the elementary steps of strand exchange can be inferred to occur on the microsecond timescale, orders of magnitude faster than the millisecond timescales of nucleotide incorporation during transcription (Roberts et al., 2008). Intriguingly, to the best of our knowledge, the observation of this mechanism within the *E. coli* SRP RNA cotranscriptional folding pathway is the first observation of toehold-mediated strand displacement in a naturally derived RNA sequence.

The high evolutionary conservation of the GU wobble base pair in many RNAs that participate in key cellular processes has been rationalized by the unique chemical and structural properties of this base pair (Varani and McClain, 2000). As exemplified in this study, the context-dependent, conformational “softness” of the H1 GU base pair may additionally allow it to act as a tripwire that triggers structural transitions of non-native to native states. Interestingly, U21 in this study is conserved in small bacterial SRP RNAs (Kalvari et al., 2018), and follow-up studies could assess if conservation of other GU base-pair structural intermediates are important for cotranscriptional rearrangement.

While this article was being prepared, an independent study of the cotranscriptional folding pathway of the same *E. coli* SRP

RNA sequence was performed using single-molecule optical tweezers (Fukuda et al., 2020). This study revealed several structural features consistent with our findings, including the formation of H1, the formation of H3 (e.g., denoted as H4 in that study), and the effect of the U21C mutation (U18C in that study) on the folding pathway. Fukuda et al. also document “hopping dynamics” near the major rearrangement, consisting of large-scale fluctuations in RNA end-to-end distances. Based on our findings, these hopping dynamics could stem from the molecular search for toehold-seeding interactions or strand-displacement attempts that open structural elements before rearrangement. These two complementary studies highlight the power of combining orthogonal approaches to gain a deeper and more complete mechanistic view of cotranscriptional RNA folding.

A fascinating question in biology pertains to how RNAs efficiently fold into functional states and exit the kinetic traps imposed by the polarity and timescale of cotranscriptional folding in the cell. While a plethora of other interactions and processes *in vivo* could facilitate these structural rearrangements, many cellular RNAs may exploit SRP RNA-like rearrangement pathways to arrive at their respective final structure. Our demonstration of how a change in the identity of a single nucleotide alters the folding trajectory also hints at how simple genetic changes could have spawned new functions in the early RNA world.

Limitations

R2D2 has several limitations, some of which are inherent in the underlying algorithms used to sample possible structures. Specifically, there are currently no efficient methods to sample RNA structures with pseudoknots, non-canonical base pairs, or RNA-ligand/RNA-protein interactions (Ding et al., 2004; Tan et al., 2017). Structures that can be efficiently sampled are biased to

the equilibrium Boltzmann distribution, which we try to overcome by sampling 150,000 states at each RNA length instead of the more commonly used 1,000–10,000 (Kutchko et al., 2015; Li and Aviran, 2018; Ouyang et al., 2013) (Figure S1). While all-atom MD was used to connect selected secondary structural states, it is inefficient to connect all possible sets of states together to reconstruct a full dynamic cotranscriptional folding pathway.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **METHOD DETAILS**
 - Cotranscriptional and equilibrium-refolded SHAPE-seq
 - RNase P assay
 - Reactivity calculation
 - DUETT
 - Reconstructing RNA secondary structures
 - Benchmarking
 - Application to cotranscriptional SHAPE-seq data
 - Minimum free energy folding pathway prediction
 - KineFold predictions
 - Using R2D2 predictions to inform all-atom folding pathway simulations
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Additional resources

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2020.12.017>.

ACKNOWLEDGMENTS

We thank Eric Siggia (Rockefeller University) and Dave Matthews (University of Rochester) for foundational conversations, Daniel Aalberts (Williams College) for assistance with plotting RNAAbow plots, Gregory Phillips and Nicholas Backes (Iowa State University) for discussions and protocols, Katherine Berman for cloning SRP RNA sequence mutants, Eric Strobel and Kyle Watters for help with cotranscriptional SHAPE-seq, and Molly Evans for SHAPE-seq protocol discussions and improvements. We also thank Shannon Yan, Shingo Fukuda, and Carlos Bustamante (University of California, Berkeley) for helpful discussions. Support for this work was provided by the Tri-Institutional Training Program in Computational Biology and Medicine (via National Institutes of Health [NIH] training grant T32GM083937 to A.M.Y., NIH grant 1DP2GM110838 to J.B.L., NIH grant GM120582 to V.G. and J.B.L., National Science Foundation [NSF] grant MCB1651877 to A.A.C., and NSF grant 1914567 to J.B.L.), and Searle Funds at The Chicago Community Trust (J.B.L.). This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (allocation TG-MCB140273 to A.A.C.), which is supported by NSF grant ACI-1548562, as well as the Quest high performance computing facility at Northwestern University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

AUTHOR CONTRIBUTIONS

Conceptualization, A.M.Y., P.M.G., A.A.C., and J.B.L.; Methodology, A.M.Y., P.M.G., L.C., L.B.L., V.G., J.B.L., and A.A.C.; Software, A.M.Y., P.M.G., J.B.L., and A.A.C.; Validation, L.C.; Formal Analysis, A.M.Y., P.M.G., and S.K.; Investigation, A.M.Y., P.M.G., L.C., L.B.L., and S.K.; Writing – Original Draft, A.M.Y., P.M.G., L.B.L., V.G., A.A.C., and J.B.L.; Writing – Review & Editing, A.M.Y., P.M.G., L.C., L.B.L., S.K., V.G., A.A.C., and J.B.L.; Visualization, A.M.Y., P.M.G., L.B.L., S.K., and A.A.C.; Supervision, V.G., A.A.C., and J.B.L.; Funding Acquisition, V.G., A.A.C., and J.B.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 29, 2020

Revised: December 8, 2020

Accepted: December 10, 2020

Published: January 14, 2021

REFERENCES

- Aalberts, D.P., and Jannen, W.K. (2013). Visualizing RNA base-pairing probabilities with RNAAbow diagrams. *RNA* 19, 475–478.
- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25.
- Al-Hashimi, H.M., and Walter, N.G. (2008). RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* 18, 321–329.
- Altman, S. (2007). A view of RNase P. *Molecular Biosyst.* 3, 604–607.
- Aviran, S., Trapnell, C., Lucks, J.B., Mortimer, S.A., Luo, S., Schroth, G.P., Doudna, J.A., Arkin, A.P., and Pachter, L. (2011). Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. USA* 108, 11069–11074.
- Batey, R.T., Rambo, R.P., Lucast, L., Rha, B., and Doudna, J.A. (2000). Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* 287, 1232–1239.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690.
- Bindewald, E., Wendeler, M., Legiewicz, M., Bona, M.K., Wang, Y., Pritt, M.J., Le Grice, S.F.J., and Shapiro, B.A. (2011). Correlating SHAPE signatures with three-dimensional RNA structures. *RNA* 17, 1688–1696.
- Bothwell, A.L., Garber, R.L., and Altman, S. (1976). Nucleotide sequence and in vitro processing of a precursor molecule to *Escherichia coli* 4.5 S RNA. *J. Biol. Chem.* 251, 7709–7716.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101.
- Chen, S.-J., and Dill, K.A. (2000). RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA* 97, 646–651.
- Chen, A.A., and Garcia, A.E. (2013). High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 110, 16820–16825.
- Danilova, L.V., Pervouchine, D.D., Favorov, A.V., and Mironov, A.A. (2006). RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J. Bioinform. Comput. Biol.* 4, 589–596.
- Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* 106, 97–102.
- Ding, Y., Chan, C.Y., and Lawrence, C.E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* 32, W135–141.
- Frieda, K.L., and Block, S.M. (2012). Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* 338, 397–400.

- Fukuda, S., Yan, S., Komi, Y., Sun, M., Gabizon, R., and Bustamante, C. (2020). The biogenesis of SRP RNA is modulated by an RNA folding intermediate attained during transcription. *Mol. Cell* 77, 241–250.e8.
- Geis, M., Flamm, C., Wolfinger, M.T., Tanzer, A., Hofacker, I.L., Middendorf, M., Mandl, C., Stadler, P.F., and Thurner, C. (2008). Folding kinetics of large RNAs. *J. Mol. Biol.* 379, 160–173.
- Gopalan, V., Baxevasis, A.D., Landsman, D., and Altman, S. (1997). Analysis of the functional role of conserved residues in the protein subunit of ribonuclease P from *Escherichia coli*. *J. Mol. Biol.* 267, 818–829.
- Heilman-Miller, S.L., and Woodson, S.A. (2003). Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA* 9, 722–733.
- Hofacker, I.L., Flamm, C., Heine, C., Wolfinger, M.T., Scheuermann, G., and Stadler, P.F. (2010). BarMap: RNA folding on dynamic energy landscapes. *RNA* 16, 1308–1316.
- Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L., and Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120, 9665–9678.
- Hsu, L.M., Zagorski, J., and Fournier, M.J. (1984). Cloning and sequence analysis of the *Escherichia coli* 4.5 S RNA gene. *J. Mol. Biol.* 178, 509–531.
- Jomaa, A., Fu, Y.-H.H., Boehringer, D., Leibundgut, M., Shan, S.O., and Ban, N. (2017). Structure of the quaternary complex between SRP, SR, and translocon bound to the translating ribosome. *Nat. Commun.* 8, 15470.
- Joung, I.S., and Cheatham, T.E., 3rd (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112, 9020–9041.
- Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46 (D1), D335–D342.
- Kim, H., Abeyirigunawardena, S.C., Chen, K., Mayerle, M., Ragunathan, K., Luthey-Schulten, Z., Ha, T., and Woodson, S.A. (2014). Protein-guided RNA dynamics during early ribosome assembly. *Nature* 506, 334–338.
- Kolb, F.A., Malmgren, C., Westhof, E., Ehresmann, C., Ehresmann, B., Wagner, E.G., and Romby, P. (2000). An unusual structure formed by anti-sense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment. *RNA* 6, 311–324.
- Komissarova, N., and Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. USA* 95, 14699–14704.
- Kramer, F.R., and Mills, D.R. (1981). Secondary structure formation during RNA synthesis. *Nucleic Acids Res.* 9, 5109–5124.
- Kutchko, K.M., Sanders, W., Ziehr, B., Phillips, G., Solem, A., Halvorsen, M., Weeks, K.M., Moorman, N., and Laederach, A. (2015). Multiple conformations are a conserved and regulatory feature of the RB1 5' UTR. *RNA* 21, 1274–1285.
- Li, H., and Aviran, S. (2018). Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* 9, 606.
- Loughrey, D., Watters, K.E., Settle, A.H., and Lucks, J.B. (2014). SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* 42, 42.
- Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA duplex map in living cells reveals higher order transcriptome structure. *Cell* 165, 1267–1279.
- McGinnis, J.L., Dunkle, J.A., Cate, J.H., and Weeks, K.M. (2012). The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.* 134, 6617–6624.
- Mortimer, S.A., and Weeks, K.M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129, 4144–4145.
- Mustoe, A.M., Brooks, C.L., and Al-Hashimi, H.M. (2014). Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* 83, 441–466.
- Ouyang, Z., Snyder, M.P., and Chang, H.Y. (2013). SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* 23, 377–387.
- Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004). Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* 2, 461–472.
- Pan, T., and Sosnick, T. (2006). RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* 35, 161–175.
- Pan, T., Artsimovitch, I., Fang, X.W., Landick, R., and Sosnick, T.R. (1999). Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. USA* 96, 9545–9550.
- Peck-Miller, K.A., and Altman, S. (1991). Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* 221, 1–5.
- Proctor, J.R., and Meyer, I.M. (2013). COFOLD: an RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Res.* 41, e102.
- Quarrier, S., Martin, J.S., Davis-Neulander, L., Beaugregard, A., and Laederach, A. (2010). Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA* 16, 1108–1117.
- Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129.
- Roberts, J.W., Shankar, S., and Filter, J.J. (2008). RNA polymerase elongation factors. *Annu. Rev. Microbiol.* 62, 211–233.
- Rosenblad, M.A., Larsen, N., Samuelsson, T., and Zwieb, C. (2009). Kinship in the SRP RNA family. *RNA Biol.* 6, 508–516.
- Saldi, T., Fong, N., and Bentley, D.L. (2018). Transcription elongation rate affects nascent histone pre-mRNA folding and 3' end processing. *Genes Dev.* 32, 297–308.
- Serganov, A., and Nudler, E. (2013). A decade of riboswitches. *Cell* 152, 17–24.
- Spasic, A., Assmann, S.M., Bevilacqua, P.C., and Mathews, D.H. (2018). Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* 46, 314–323.
- Steinbrecher, T., Latzer, J., and Case, D.A. (2012). Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* 8, 4405–4412.
- Šulc, P., Ouldrige, T.E., Romano, F., Doye, J.P., and Louis, A.A. (2015). Modelling toehold-mediated RNA strand displacement. *Biophys. J.* 108, 1238–1247.
- Tan, Z., Sharma, G., and Mathews, D.H. (2017). Modeling RNA secondary structure with sequence comparison and experimental mapping data. *Biophys. J.* 113, 330–338.
- Varani, G., and McClain, W.H. (2000). The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1, 18–23.
- Vioque, A., Arnez, J., and Altman, S. (1988). Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *J. Mol. Biol.* 202, 835–848.
- Wang, J., Cieplak, P., and Kollman Peter, A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21, 1049–1074.
- Watters, K.E., Strobel, E.J., Yu, A.M., Lis, J.T., and Lucks, J.B. (2016a). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.* 23, 1124–1131.
- Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H., and Lucks, J.B. (2016b). Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* 103, 34–48.
- Whitford, P.C., Schug, A., Saunders, J., Hennelly, S.P., Onuchic, J.N., and Sanbonmatsu, K.Y. (2009). Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function. *Biophys. J.* 96, L7–L9.

- Wiebe, N.J.P., and Meyer, I.M. (2010). TRANSAT—method for detecting the conserved helices of functional RNA structures, including transient, pseudo-knotted and alternative structures. *PLoS Comput. Biol.* *6*, e1000823.
- Wong, T.N., Sosnick, T.R., and Pan, T. (2007). Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc. Natl. Acad. Sci. USA* *104*, 17995–18000.
- Xayaphoummine, A., Bucher, T., and Isambert, H. (2005). Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* *33*, W605–10.
- Xue, A.Y., Yu, A.M., Lucks, J.B., and Bagheri, N. (2019). DUETT quantitatively identifies known and novel events in nascent RNA structural dynamics from chemical probing data. *Bioinformatics* *35*, 5103–5112.
- Zhang, D.Y., and Winfree, E. (2009). Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* *131*, 17303–17314.
- Zhu, J.Y.A., Steif, A., Proctor, J.R., and Meyer, I.M. (2013). Transient RNA structure features are evolutionarily conserved and can be computationally predicted. *Nucleic Acids Res.* *41*, 6273–6285.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
NEB Turbo Competent <i>E. coli</i> (High Efficiency)	New England Biolabs	Cat#C2984H
Biological samples		
NEB Turbo Competent <i>E. coli</i> (High Efficiency)	New England Biolabs	Cat#C2984H
Chemicals, peptides, and recombinant proteins		
Vent (exo-) DNA Polymerase	New England Biolabs	Cat#M0257S
Deoxynucleotide (dNTP) Solution Mix	New England Biolabs	Cat#N9447L
<i>E. coli</i> RNA Polymerase, Holoenzyme	New England Biolabs	Cat#M0551S
Gln111 (EcoRI E111Q Mutant)	Lab preparation	N/A
Ribonucleotide Solution Set	New England Biolabs	Cat#N0466
Benzoyl Cyanide	Sigma-Aldrich	Cat#115959
Dimethyl Sulfoxide	Sigma-Aldrich	Cat#276855
TRIzol Reagent	Thermo Fisher	Cat#15596018
Isopropyl alcohol	Sigma-Aldrich	Cat#I9516
Chloroform	Sigma-Aldrich	Cat#C2432
Sodium Acetate	Sigma-Aldrich	Cat#S2889
Glycogen, RNA grade	Thermo Fisher	Cat#R0551
Ethyl alcohol, pure	Sigma-Aldrich	E7023
SuperScript III Reverse Transcriptase	Thermo Fisher	Cat#18080093
CircLigase ssDNA Ligase	Lucigen	Cat#CL4111K
Phusion High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0530L
GeneScan 500 LIZ Size Standard	Applied Biosystems	Cat#4322682
<i>E. coli</i> C5 protein	Gopalan et al., 1997	pBSC5
T7 RNA polymerase	Lab preparation	pQE9T7
ATP	Carbosynth	Cat# NA00135
CTP	Carbosynth	Cat# NC03860
GTP	Carbosynth	Cat# NG01208
UTP	Carbosynth	Cat# NU03863
Deposited data		
Signal Recognition Particle RNA cotranscriptional SHAPE-Seq, Replicate 1	Watters et al., 2016a	SRX2159310
Signal Recognition Particle RNA cotranscriptional SHAPE-Seq, Replicate 2	Watters et al., 2016a	SRX2159311
Signal Recognition Particle RNA cotranscriptional SHAPE-Seq, Replicate 3	Watters et al., 2016a	SRX2159312
Signal Recognition Particle RNA equilibrium-refolded SHAPE-Seq	Watters et al., 2016a	SRX2159316
U21C Signal Recognition Particle RNA cotranscriptional SHAPE-Seq	This paper	PRJNA667733
U21C Signal Recognition Particle RNA equilibrium-refolded SHAPE-Seq	This paper	PRJNA667733
U21C, C22U, G93A Signal Recognition Particle RNA cotranscriptional SHAPE-Seq	This paper	PRJNA667733
U21C Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 1	This paper	SRPU21C_BZCN_0001

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
U21C Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 2	This paper	SRPU21C_BZCN_0002
U21C Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 3	This paper	SRPU21C_BZCN_0003
U21C Signal Recognition Particle RNA equilibrium-refolded SHAPE-Seq reactivities, Replicate 1	This paper	SRPU21C_BZCN_0004
U21C Signal Recognition Particle RNA equilibrium-refolded SHAPE-Seq reactivities, Replicate 2	This paper	SRPU21C_BZCN_0005
U21C, C22U, G93A Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 1	This paper	SRP21CR_BZCN_0001
U21C, C22U, G93A Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 2	This paper	SRP21CR_BZCN_0002
U21C, C22U, G93A Signal Recognition Particle RNA cotranscriptional SHAPE-Seq reactivities, Replicate 3	This paper	SRP21CR_BZCN_0003

Oligonucleotides

ttttttgaattcGACCTGACCTGGTAAACAGA	IDT	EJS_F01_wt_ER1_55.R
ttttttgaattcGGACCTGACCTGGTAAACAG	IDT	EJS_F02_wt_ER1_56.R
ttttttgaattcCGGACCTGACCTGGTAAACA	IDT	EJS_F03_wt_ER1_57.R
ttttttgaattcCCGGACCTGACCTGGTAAAC	IDT	EJS_F04_wt_ER1_58.R
ttttttgaattcTCCGGACCTGACCTGGTAAA	IDT	EJS_F05_wt_ER1_59.R
ttttttgaattcTCCGGACCTGACCTGGTAA	IDT	EJS_F06_wt_ER1_60.R
ttttttgaattcCTCCGGACCTGACCTGGTA	IDT	EJS_F07_wt_ER1_61.R
ttttttgaattcCCTCCGGACCTGACCTGGT	IDT	EJS_F08_wt_ER1_62.R
ttttttgaattcTCCTCCGGACCTGACCTGG	IDT	EJS_F09_wt_ER1_63.R
ttttttgaattcTTCCCTCCGGACCTGACCTG	IDT	EJS_F10_wt_ER1_64.R
ttttttgaattcCTTCCCTCCGGACCTGACCT	IDT	EJS_F11_wt_ER1_65.R
ttttttgaattcGCTTCCCTCCGGACCTGACC	IDT	EJS_F12_wt_ER1_66.R
ttttttgaattcTGCTTCCCTCCGGACCTGAC	IDT	EJS_F13_wt_ER1_67.R
ttttttgaattcCTGCTTCCCTCCGGACCTGA	IDT	EJS_F14_wt_ER1_68.R
ttttttgaattcGCTGCTTCCCTCCGGACCTG	IDT	EJS_F15_wt_ER1_69.R
ttttttgaattcGGCTGCTTCCCTCCGGACCT	IDT	EJS_F16_wt_ER1_70.R
ttttttgaattcTGGCTGCTTCCCTCCGGACC	IDT	EJS_F17_wt_ER1_71.R
ttttttgaattcTTGGCTGCTTCCCTCCGGAC	IDT	EJS_F18_wt_ER1_72.R
ttttttgaattcCTTGGCTGCTTCCCTCCGGA	IDT	EJS_F19_wt_ER1_73.R
ttttttgaattcCCTTGGCTGCTTCCCTCCGG	IDT	EJS_F20_wt_ER1_74.R
ttttttgaattcGCCTTGGCTGCTTCCCTCCG	IDT	EJS_F21_wt_ER1_75.R
ttttttgaattcTGCCCTGGCTGCTTCCCTTCC	IDT	EJS_F22_wt_ER1_76.R
ttttttgaattcCTGCCCTGGCTGCTTCCCTTCC	IDT	EJS_F23_wt_ER1_77.R
ttttttgaattcTCTGCCCTGGCTGCTTCCCTT	IDT	EJS_F24_wt_ER1_78.R
ttttttgaattcATCTGCCCTGGCTGCTTCCCT	IDT	EJS_F25_wt_ER1_79.R
ttttttgaattcCATCTGCCCTGGCTGCTTCC	IDT	EJS_F26_wt_ER1_80.R
ttttttgaattcTCATCTGCCCTGGCTGCTTCC	IDT	EJS_F27_wt_ER1_81.R
ttttttgaattcGTCATCTGCCCTGGCTGCTT	IDT	EJS_F28_wt_ER1_82.R
ttttttgaattcCGTCATCTGCCCTGGCTGCT	IDT	EJS_F29_wt_ER1_83.R
ttttttgaattcGCGTCATCTGCCCTGGCTGC	IDT	EJS_F30_wt_ER1_84.R
ttttttgaattcCGCGTCATCTGCCCTGGCTG	IDT	EJS_F31_wt_ER1_85.R
ttttttgaattcaCGCGTCATCTGCCCTGGCT	IDT	EJS_F32_wt_ER1_86.R
ttttttgaattcCaCGCGTCATCTGCCCTGGC	IDT	EJS_F33_wt_ER1_87.R
ttttttgaattcaCaCGCGTCATCTGCCCTGG	IDT	EJS_F34_wt_ER1_88.R

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ttttttgaattcCaCaCGCGTCATCTGCCTTG	IDT	EJS_F35_wt_ER1_89.R
ttttttgaattcGCaCaCGCGTCATCTGCCTT	IDT	EJS_F36_wt_ER1_90.R
ttttttgaattcGGCaCaCGCGTCATCTGCCT	IDT	EJS_F37_wt_ER1_91.R
ttttttgaattcCGGCaCaCGCGTCATCTGCC	IDT	EJS_F38_wt_ER1_92.R
ttttttgaattcCCGGCaCaCGCGTCATCTGC	IDT	EJS_F39_wt_ER1_93.R
ttttttgaattcCCCGCaCaCGCGTCATCTG	IDT	EJS_F40_wt_ER1_94.R
ttttttgaattcTCCCGCaCaCGCGTCATCT	IDT	EJS_F41_wt_ER1_95.R
ttttttgaattcATCCCGCaCaCGCGTCATC	IDT	EJS_F42_wt_ER1_96.R
ttttttgaattcCATCCCGCaCaCGCGTCAT	IDT	EJS_F43_wt_ER1_97.R
ttttttgaattcACATCCCGCaCaCGCGTCA	IDT	EJS_F44_wt_ER1_98.R
ttttttgaattcTACATCCCGCaCaCGCGTC	IDT	EJS_F45_wt_ER1_99.R
ttttttgaattcCTACATCCCGCaCaCGCGT	IDT	EJS_F46_wt_ER1_100.R
ttttttgaattcGCTACATCCCGCaCaCGCG	IDT	EJS_F47_wt_ER1_101.R
ttttttgaattcAGCTACATCCCGCaCaCGC	IDT	EJS_F48_wt_ER1_102.R
ttttttgaattcCAGCTACATCCCGCaCaCG	IDT	EJS_F49_wt_ER1_103.R
ttttttgaattcCCAGCTACATCCCGCaCaC	IDT	EJS_F50_wt_ER1_104.R
ttttttgaattcGCCAGCTACATCCCGCaCa	IDT	EJS_F51_wt_ER1_105.R
ttttttgaattcTGCCAGCTACATCCCGCaC	IDT	EJS_F52_wt_ER1_106.R
ttttttgaattcCTGCCAGCTACATCCCGCa	IDT	EJS_F53_wt_ER1_107.R
ttttttgaattcCCTGCCAGCTACATCCCGGC	IDT	EJS_F54_wt_ER1_108.R
ttttttgaattcCCCTGCCAGCTACATCCCGG	IDT	EJS_F55_wt_ER1_109.R
ttttttgaattcGCCCTGCCAGCTACATCCCG	IDT	EJS_F56_wt_ER1_110.R
ttttttgaattcGGCCCTGCCAGCTACATCCC	IDT	EJS_F57_wt_ER1_111.R
ttttttgaattcGGGCCCTGCCAGCTACATCC	IDT	EJS_F58_wt_ER1_112.R
ttttttgaattcGGGGCCCTGCCAGCTACATC	IDT	EJS_F59_wt_ER1_113.R
ttttttgaattcGGGGGCCCTGCCAGCTACAT	IDT	EJS_F60_wt_ER1_114.R
ttttttgaattcTGGGGGCCCTGCCAGCTACA	IDT	EJS_F61_wt_ER1_115.R
ttttttgaattcGTGGGGGCCCTGCCAGCTAC	IDT	EJS_F62_wt_ER1_116.R
ttttttgaattcGGTGGGGGCCCTGCCAGCTA	IDT	EJS_F63_wt_ER1_117.R
ttttttgaattcGGGTGGGGGCCCTGCCAGCT	IDT	EJS_F64_wt_ER1_118.R
ttttttgaattcCGGGTGGGGGCCCTGCCAGC	IDT	EJS_F65_wt_ER1_119.R
ttttttgaattcCCGGTGGGGGCCCTGCCAG	IDT	EJS_F66_wt_ER1_120.R
ttttttgaattcCCCGGTGGGGGCCCTGCCA	IDT	EJS_F67_wt_ER1_121.R
ttttttgaattcACCCGGTGGGGGCCCTGCC	IDT	EJS_F68_wt_ER1_122.R
ttttttgaattcGACCCGGTGGGGGCCCTGC	IDT	EJS_F69_wt_ER1_123.R
ttttttgaattcCGACCCGGTGGGGGCCCTG	IDT	EJS_F70_wt_ER1_124.R
ttttttgaattcCCGACCCGGTGGGGGCCCT	IDT	EJS_F71_wt_ER1_125.R
ttttttgaattcGCCGACCCGGTGGGGGCC	IDT	EJS_F72_wt_ER1_126.R
ttttttgaattcTGCCGACCCGGTGGGGGCC	IDT	EJS_F73_wt_ER1_127.R
ttttttgaattcATGCCGACCCGGTGGGGGC	IDT	EJS_F74_wt_ER1_128.R
ttttttgaattcCATGCCGACCCGGTGGGGG	IDT	EJS_F75_wt_ER1_129.R
ttttttgaattcCCATGCCGACCCGGTGGGG	IDT	EJS_F76_wt_ER1_130.R
ttttttgaattcGCCATGCCGACCCGGTGGG	IDT	EJS_F77_wt_ER1_131.R
ttttttgaattcTGCCATGCCGACCCGGTGG	IDT	EJS_F78_wt_ER1_132.R
ttttttgaattcATGCCATGCCGACCCGGTGG	IDT	EJS_F79_wt_ER1_133.R
ttttttgaattcGATGCCATGCCGACCCGGT	IDT	EJS_F80_wt_ER1_134.R
ttttttgaattcAGATGCCATGCCGACCCGGG	IDT	EJS_F81_wt_ER1_135.R
ttttttgaattcGAGATGCCATGCCGACCCGG	IDT	EJS_F82_wt_ER1_136.R

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ttttttgaattcGGAGATGCCATGCCGACCCG	IDT	EJS_F83_wt_ER1_137.R
ttttttgaattcTGGAGATGCCATGCCGACCC	IDT	EJS_F84_wt_ER1_138.R
ttttttgaattcGTGGAGATGCCATGCCGACC	IDT	EJS_F85_wt_ER1_139.R
ttttttgaattcGGTGGAGATGCCATGCCGAC	IDT	EJS_F86_wt_ER1_140.R
ttttttgaattcAGGTGGAGATGCCATGCCGA	IDT	EJS_F87_wt_ER1_141.R
ttttttgaattcGAGGTGGAGATGCCATGCCG	IDT	EJS_F88_wt_ER1_142.R
ttttttgaattcGGAGGTGGAGATGCCATGCC	IDT	EJS_F89_wt_ER1_143.R
ttttttgaattcAGGAGGTGGAGATGCCATGC	IDT	EJS_F90_wt_ER1_144.R
ttttttgaattcGAGGAGGTGGAGATGCCATG	IDT	EJS_F91_wt_ER1_145.R
ttttttgaattcCGAGGAGGTGGAGATGCCAT	IDT	EJS_F92_wt_ER1_146.R
ttttttgaattcGCGAGGAGGTGGAGATGCCA	IDT	EJS_F93_wt_ER1_147.R
ttttttgaattcCGCGAGGAGGTGGAGATGCC	IDT	EJS_F94_wt_ER1_148.R
ttttttgaattcCCGCGAGGAGGTGGAGATGC	IDT	EJS_F95_wt_ER1_149.R
ttttttgaattcCCGCGAGGAGGTGGAGATGC	IDT	EJS_F96_wt_ER1_150.R
ttttttgaattcACCGCGAGGAGGTGGAGATG	IDT	EJS_F01_wt_ER1_55.R
/5Phos/rCrUrGrArCrUrCrGrGrCrArCrCrArArGrGrA/3ddC/	IDT	Linker
/5Biosg/gtccttggtgcccgagt	IDT	RT Primer
/5Phos/AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC/3SpC3/	IDT	A_Adapter_B
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	IDT	PE_Forward
CTTTCCCTACACGACGCTCTTCCGATCT	IDT	Select (+)
TCCGATCTRRRYGTCTTG GTGCCCGAG*T*c*a*g		
CTTTCCCTACACGACGCTCTTCCGATCTCTTCCGATCTYYYRGTCCTTGGTGCCCGAG*T*c*a*g	IDT	Select (-)
CAAGCAGAAGACGGCATACGAGATN NNNNNGTGACTGGAGTTCAGACGTGTGCTC	IDT	Illumina Index
GAGCGCGCGTAATACGACTCACTATAGCGT TGGTTCTCAACGCTCTCAATG	Sigma-Aldrich	4.5S-F
TGCGGGAGAACCAACAGAGCC CCATTGAGAGCGTTGAGAAC	Sigma-Aldrich	4.5S-R
CCTTGGCTGCTTCCCTCCGG	Sigma-Aldrich	4.5S(70)-R
Recombinant DNA		
pJBL3664_SRP_EcoliRNAP_trp_HepD	Watters et al., 2016a	RRID:Addgene_162240
p23-4.5S	Peck-Miller and Altman, 1991	N/A
Software and algorithms		
R2D2	This paper	https://github.com/LucksLab/R2D2
Cotranscriptional SHAPE-Seq Tools	Watters et al., 2016a	https://github.com/LucksLab/Cotrans_SHAPE-Seq_Tools
RNAstructure	Reuter and Mathews, 2010	https://rna.urmc.rochester.edu/RNAstructure.html
KineFold	Xayaphoummine et al., 2005	http://kinefold.curie.fr/download.html
GROMACS 2016	Abraham et al., 2015	http://ftp.gromacs.org/pub/gromacs/gromacs-2016.6.tar.gz
Other		
M1 RNA	Vioque et al., 1988	pJA2'
Detailed Protocol	This paper	Methods S1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Julius Lucks (jlucks@northwestern.edu).

Materials availability

The *E. coli* SRP RNA plasmid used in this study will be available through Addgene ID 162240.

Data and code availability

The accession number for the mutant cotranscriptional and equilibrium-refolded SHAPE-seq datasets reported in this paper is SRA: PRJNA667733. Wild-type SRP RNA data are similarly available through the Small Read Archive (SRA) with accession codes: SRA:SRX2159310, SRX2159311, SRX2159312, and SRX2159316. Processed SHAPE-seq reactivity files generated in this study will be deposited in the RNA Mapping Database (RMDb) under accession codes RMDb: SRPU21C_BZCN_0001, SRPU21C_BZCN_0002, SRPU21C_BZCN_0003, SRPU21C_BZCN_0004, SRPU21C_BZCN_0005, SRP21CR_BZCN_0001, SRP21CR_BZCN_0002, and SRP21CR_BZCN_0003. All source code will be freely available at <https://github.com/LucksLab/R2D2>. For a single round 2D folding pathway prediction with 2 processors, the walltime used is around 4 hours and used around 26 GB memory. We run this 100 times before analyzing all 100 results. The 3D folding simulations of the SRP precursor went for 48 ns using 9 synchronous replicas at different temperatures and restraint strengths. Each replica took 142 hours (–5.9 days) using an entire STAMPEDE-2 Intel Xeon Phi 7250 node. Thus, the experiment took altogether 1,275 node-hours (or 53 node-days) for the one folding simulation. Each of the other pathways examined used comparable resources.

METHOD DETAILS

Cotranscriptional and equilibrium-refolded SHAPE-seq

The SRP RNA sequence used to generate mutants was previously described ([Watters et al., 2016a](#)); it has an AUC sequence substituting the 24-nt leader. DNA templates for cotranscriptional SHAPE-seq were prepared as previously described ([Watters et al., 2016a](#)). DNA templates specifically targeted transcript lengths 101 to 136 for U21C and U21C, C22U, G93A mutants. Cotranscriptional SHAPE-seq experiments were performed as previously described, except that EcoRI_{E111Q} was included at 800 nM during *in vitro* transcription instead of 500 nM.

RNase P assay

The 24+24 pre-SRP RNA used to examine co-transcriptional processing was generated by run-off *in vitro* transcription (IVT). The template for this IVT was obtained by annealing two overlapping DNA oligonucleotides 4.5S-F and 4.5S-R (Sigma-Aldrich) followed by filling-in with Phusion DNA polymerase (NEB) to obtain a double-stranded DNA that included a T7 promoter upstream of the RNA coding sequence. A portion of the transcribed 24+24 pre-SRP RNA was 5'-radiolabeled by dephosphorylating with calf intestinal phosphatase (NEB) and then phosphorylated with [γ -³²P]-ATP (PerkinElmer) using polynucleotide kinase (NEB). To determine the cleavage efficiency of the 24+24 pre-SRP RNA by RNase P, *E. coli* RNase P was reconstituted *in vitro* using recombinant M1 RNA and C5 protein ([Gopalan et al., 1997](#)). *In vitro* transcribed M1 RNA (2 μ M) was refolded in water at 50°C for 50 min, then 37°C for 10 min. An equal volume of 2x folding buffer [20 mM HEPES (pH 7.5), 800 mM NH₄OAc, 20 mM Mg(OAc)₂, 10% glycerol, 0.02% IGEPAL] was added, and incubation at 37°C was continued for 30 min. C5 protein was overexpressed and purified from *E. coli* as described previously ([Vioque et al., 1988](#)) and was stored at –80°C. Before use, the refolded M1 RNA and C5 protein were diluted to 0.1 and 1 μ M, respectively, in assay buffer [1x = 20 mM Tris-HCl (pH 8), 50 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 0.2 mg/mL BSA, 1 mM DTT]. This assay buffer resembles the one used in cotranscriptional folding experiments to mimic the same condition. All following incubations were performed at 37°C in a thermal cycler. For each 20 μ L reaction, a mixture containing 6 μ L water, 8 μ L 2x assay buffer, and 2 μ L of 1 μ M C5 protein (final 100 nM) was incubated for 5 min before adding 2 μ L of 0.1 μ M M1 RNA (final 10 nM) and continuing incubation for 10 min. The reaction was initiated by adding 2 μ L of the 24+24 pre-SRP RNA, where the final concentration (10 – 2,500 nM) was made up of the unlabeled RNA and a trace amount of the radiolabeled RNA. After each specified time interval, a 3 μ L aliquot of the reaction was removed and quenched with 10 μ L termination dye [7 M urea, 1 mM EDTA, 0.05% (w/v) each of bromophenol blue and xylene cyanol, 10% (v/v) phenol]. The products and uncleaved substrate were then separated on an 8% (w/v) polyacrylamide/7 M urea gel. The gels were visualized by phosphorimaging on the Typhoon (GE Healthcare), and bands were quantitated using ImageQuant (GE Healthcare). As ladders to map the cleavage site, a 24+70 pre-SRP RNA was generated by IVT with a template that was PCR-amplified from p23-4.5S ([Peck-Miller and Altman, 1991](#)) using primers 4.5S-F and 4.5S(70)-R. This RNA was then 5'-radiolabeled as described above and used to make the alkaline hydrolysis ladder and the RNase T1 (Invitrogen)-generated G-ladder.

Reactivity calculation

Quantification of reactivities from cotranscriptional SHAPE-Seq data was performed using Spats v.1.0.1 (<http://luckslab.github.io/spats/>) as previously described (Watters et al., 2016a). The θ reactivities output by Spats were converted to ρ reactivities to allow for direct comparison of SHAPE probe accessibility between intermediate lengths of RNAs (Watters et al., 2016b). For cotranscriptional predictions where RNA polymerase occludes the last ~ 14 nts from folding (Komissarova and Kashlev, 1998; Watters et al., 2016a), ρ reactivities were trimmed by 14 nts and renormalized such that the reactivities average to 1. This trimming was not done for equilibrium-refolded predictions because the RNAs have already emerged from the RNA polymerase.

DUETT

Detection of Unknown Events with Tunable Thresholds (DUETT) was used to detect reactivity change events in cotranscriptional and equilibrium-refolded SHAPE-Seq datasets (Xue et al., 2019). All analyses were done with optimized parameters with window sizes of 4 for U21C and 9 for U21C, C22U, G93A.

Reconstructing RNA secondary structures

The R2D2 sample-and-select method was first developed to predict the equilibrium fold of a single RNA using equilibrium SHAPE-Seq data. A crucial step was to establish a method to *select* structures that are most consistent with the experiment. SHAPE-Seq reactivities, ρ , are values ≥ 0 that reflect the structural state of each nucleotide: $\rho = 0$ corresponds to a nucleotide that is present in a structured context (such as a base pair or stacking interaction), while $\rho > 1$ represents a nucleotide that is present in a flexible context (such as an unpaired region) (Bindewald et al., 2011). Thus ρ values most naturally correspond to a representation of the unpaired state of each nucleotide in an RNA secondary structure, which can be represented by a binary vector (u for ‘un-paired’) containing 0 if a nucleotide is paired and 1 if a nucleotide is un-paired (Figure 1C). Comparison between the ρ vector ($\vec{\rho}$) of reactivity data at a specific transcript length, and the u vector (\vec{u}) for a specific structure that could occur at that length can then be made with a metric that reflects their distance from each other (Figure 1C; Table S1).

We developed and tested six functions to calculate the distance between a SHAPE-seq reactivity spectra and a given RNA secondary structure (Figure 1C; Table S1). Each distance function is of the form

$$D_{\{K, U, D\}}^{\{cap, nocap\}}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} |\tilde{u}_i - \tilde{\rho}_i| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} |\tilde{u}_i - \tilde{\rho}_i|$$

where \tilde{u} is a vector calculated from the u -vector of a specific RNA secondary structure, and $\tilde{\rho}$ is calculated from the experimental SHAPE-seq reactivity data ρ vector. Reactivity is inherently a measure of accessibility of chemical probes in RNA structures, and low reactivity may not be due only to base pairing, but can be caused by other structural constraints such as stacking (Bindewald et al., 2011). To account for this possibility, we incorporated a weighting between single-stranded and paired bases in sampled structures, α , which is used to adjust the contribution to the distance from positions that are predicted to be paired.

Since unpaired vectors and ρ vectors are different types of data (binary versus continuous) and on different numerical scales, we explored three different ways to calculate their differences specified by K, U and D, which specify the way \tilde{u}_i and $\tilde{\rho}_i$ are calculated:

$$K: \tilde{u}_i = u_i$$

$$U: \tilde{u}_i = u_i * \text{length}(\vec{u}) / \text{sum}(\vec{u})$$

$$D: \tilde{\rho}_i = \rho_i / \max(\vec{\rho})$$

K keeps the scale of \vec{u} and $\vec{\rho}$, U makes \vec{u} 's average 1 which is a property of $\vec{\rho}$, and D scales $\vec{\rho}$ to be between 0 and 1. Since certain RNA folds can result in ρ values that are much larger than one (McGinnis et al., 2012), we also explored ways to cutoff ρ values at a maximum value. This is specified by the indices *cap* or *nocap* which determine the way $\tilde{\rho}_i$ is calculated, with *cap* denoting that ρ_i is capped at a ρ_{max} value $\tilde{\rho}_i = \min(\rho_i, \rho_{max})$, and *nocap* referring that the original ρ_i value is used. The full definitions are as follows:

$$D_K^{cap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} |u_i - \min(\rho_i, \rho_{max})| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} |u_i - \min(\rho_i, \rho_{max})|$$

$$D_K^{nocap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} |u_i - \rho_i| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} |u_i - \rho_i|$$

$$D_U^{cap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} \left| u_i * \text{length}(\vec{u}) / \text{sum}(\vec{u}) - \min(\rho_i, \rho_{max}) \right| \\ + (1 - \alpha) \sum_{i \in \text{unpaired bases}} \left| u_i * \text{length}(\vec{u}) / \text{sum}(\vec{u}) - \min(\rho_i, \rho_{max}) \right|$$

$$D_U^{nocap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} \left| u_i * \text{length}(\vec{u}) / \text{sum}(\vec{u}) - \rho_i \right| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} \left| u_i * \text{length}(\vec{u}) / \text{sum}(\vec{u}) - \rho_i \right|$$

$$D_D^{cap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} \left| u_i - \frac{\min(\rho_i, \rho_{max})}{\max(\vec{\rho})} \right| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} \left| u_i - \frac{\min(\rho_i, \rho_{max})}{\max(\vec{\rho})} \right|$$

$$D_D^{nocap}(\vec{u}, \vec{\rho}) = \alpha \sum_{i \in \text{paired bases}} \left| u_i - \frac{\rho_i}{\max(\vec{\rho})} \right| + (1 - \alpha) \sum_{i \in \text{unpaired bases}} \left| u_i - \frac{\rho_i}{\max(\vec{\rho})} \right|$$

The distance metrics above can be used to *select* structures from a candidate set that are most consistent with the observed experimental reactivity data by choosing the minimum distance structure(s) at every length (Figure 1C). To generate a candidate set of structures, the *sample* method statistically samples structures with a large sample size using the *partition* and *stochastic* functions of the RNAstructure suite of computational secondary structure prediction tools (Reuter and Mathews, 2010). The *partition* method takes as an input the RNA sequence and folding parameters, and uses them to calculate the secondary structure partition function for that sequence. The *stochastic* method then uses this partition function to stochastically generate RNA structures according to their equilibrium Boltzmann probabilities – i.e., lower free energy structures are generated more frequently than higher free energy structures. Thus repeated application of the *stochastic* method can generate a set of possible candidate structures the RNA molecule may sample during the experiment.

The goal of the *sample* method is to generate the greatest amount of structural diversity possible to allow more choices for the *select* method. An initial test of the degree to which the stochastic method can generate novel structures revealed that the method did not converge on exhausting the possibilities of different RNA structures even after 150,000 structures were drawn (Figure S1). This is not surprising since the free energy landscapes of RNA secondary structures are known to have a shallow density of states near the minimum free energy structure (Chen and Dill, 2000) indicating there are many possible RNA structures that are low in free energy and would be sampled frequently by the *stochastic* method. To circumvent this problem and still generate a diverse array of candidate structures without the computational burden of generating millions of structures, we employed two additional variations of the sampling procedure that used experimental SHAPE restraints to calculate a modified partition function from which we could sample. The first, called SHAPE-directed sampling, used the *partition* method's ability to incorporate SHAPE reactivities as effective free energy terms in the partition function calculation with pseudofree energy parameters $m = 1.1$ and $b = -0.3$. The second, called SHAPE-forced sampling, used a SHAPE reactivity cutoff, ρ_c , to force nucleotides with reactivities greater than this value to be single-stranded in the partition function calculation. In total, the *sample* method consisted of sampling 50,000 structures from each of these methods for a total of 150,000 structures which acted as the candidate set for the *select* method. We note that even though the *sample* method uses SHAPE reactivity data to generate part of the candidate set, these are not guaranteed to be chosen as most consistent with the data by the *select* method. Rather, they are included to increase the diversity of the candidate set.

Software implementing this method were run with Python 2.7.12 through Anaconda 2.4.1 (64-bit) and R version 3.2.2. Images and movies were made with ffmpeg version 3.1.3, ImageMagick 7.0.3-0 Q16 x86_64, and iMovie v10.1.2. Version 5.8.1 of RNAstructure was used for the *partition* and *stochastic* methods, and VARNA version 3.9 was used to visualize RNA secondary structures. See “Data and Software Availability” for location of code used in this study.

Benchmarking

Best parameter values were determined through a grid search of 10,404 parameter sets: all combinations of 0.7 to 4.1 by 0.1 for ρ_c , 0.7 to 4.1 by 0.1 for ρ_{max} , and 0 to 1 by 0.1 for α . The best parameter set(s) was determined as the parameter set(s) with the largest sum of F_1 scores (F-scores) for 18 previously published equilibrium-refolded SHAPE-Seq datasets on 6 RNAs of known crystal structures (three replicates) and no pseudoknots since RNAstructure cannot sample structures with pseudoknots (Loughrey et al., 2014). F-score is defined as follows:

$$F = 2 * \frac{\text{sensitivity} * \text{PPV}}{\text{sensitivity} + \text{PPV}}$$

$$\text{sensitivity} = \frac{\text{Number of true base pairs predicted}}{\text{Number of true base pairs}}$$

$$\text{PPV} = \frac{\text{Number of true base pairs predicted}}{\text{Number of predicted base pairs}}$$

For every parameter set, we sampled 50,000 structures for each of the three sampling methods, for a total of 150,000 structures (see “[Reconstructing RNA secondary structures](#)”). For each benchmark RNA and dataset, the minimum distance structure was calculated and F-score determined from the prediction and the known structure. The sum of F-scores across the panel of RNAs and datasets was then reported for that parameter set. If multiple minimum distance structures were found, then the average of their sum of F-scores were used to find the best parameter set. We ran the benchmarking for each of the 6 distance equations. ρ_{max} is not used when no reactivity capping is used, so only 306 parameter sets were tested in these cases.

We found two different metrics were the best performing across all distance functions: $D_{K,cap}$ with $\rho_c = 3.5$, $\rho_{max} = 1.0$ or 0.9 , and $\alpha = 0.8$ as well as $D_{D,cap}$ with $\rho_c = 3.5$, $\rho_{max} = 1.0$, and $\alpha = 0.8$. These two each had an average F-score of 86.32% for the 18 RNA datasets in the panel ([Table S1](#)). From this set, we chose as our parameter set $D_{K,cap}$ with $\rho_c = 3.5$, $\rho_{max} = 1.0$, and $\alpha = 0.8$, which gives a higher weight to paired positions in the sampled structures as expected, and matches common interpretations of ‘high’ reactivity values being greater than 1. We note that this is mathematically equivalent to $D_{D,cap}$ ’s best parameter set.

We also compared the best results from the sample-and-select method to SHAPE-restrained secondary structure predictions using the same data on the same RNA panel using the *Fold* method of RNAstructure ([Table S2](#)). In aggregate, the sample-and-select method (average F-score of 86.32%) does not perform better than RNAstructure-Fold with SHAPE restraints (average F-score of 88.95%), but does perform better than RNAstructure-Fold without SHAPE restraints (average F-score of 77.51%). Interestingly R2D2’s sample-and-select method did outperform on the *E. coli* TPP riboswitch in terms of sensitivity, PPV, and F-score for all replicates ([Table S2](#)). While the accuracy of our sample-and-select method applied to equilibrium RNA structure prediction is not overall better than the best equilibrium structure prediction algorithms given the same data, it was designed to find RNA secondary structures consistent with structural probing data from out-of-equilibrium RNA folds and thus can be used to reconstruct a complete secondary structure cotranscriptional folding pathway of an RNA.

To increase PPV for more accurate 3D simulations, R2D2 filters base pairs and reduces overall positive calls compared to RNAstructure-Fold; 2D sampling is run 100 times and only pairs that occur over 50% of the time are kept and then used in the 3D simulations. We assessed this filtering step (which we call R2D2-consensus) using the benchmark panel ([Table S2](#)). Both R2D2-consensus and RNAstructure-Fold with SHAPE perform better than RNAstructure-Fold with no SHAPE. The counts across true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) between R2D2-consensus and RNAstructure-Fold with SHAPE are statistically significant different at the 0.05 value (p value 0.001) by multivariate 2-sample E-test of equal distributions. As expected, R2D2-consensus disfavors calling positive base pairs compared to RNAstructure-Fold with SHAPE: R2D2-consensus predicts 478 base pairs across the whole panel compared to 522 base pairs by RNAstructure-Fold with SHAPE. Importantly, R2D2-consensus, which is the first filter of positive base pairs out of two in the R2D2 algorithm, has a reduced number of FP’s when compared to RNAstructure-Fold with and without SHAPE data. R2D2-consensus also has a lower standard deviation in sensitivity, PPV, and F-score compared to RNAstructure-Fold.

However, based on the Sensitivity, PPV, and F-score metrics alone, there is no statistical difference in performance in any of these three metrics between RNAstructure-Fold with SHAPE and R2D2-consensus using paired t test. Interestingly Sensitivity, PPV, and F-score are calculated based on TP, FN, and FP counts, and there is a difference in statistical significance when examining the same prediction results at different representation levels.

Software implementing this method were run with Python 2.7.11 through Anaconda 2.3.0 (64-bit). Version 5.6 beta of RNAstructure was used for the *partition* and *stochastic* methods, and VARNA version 3.9 was used to visualize RNA secondary structures. See “[Data and Software Availability](#)” for location of code used in this study.

Application to cotranscriptional SHAPE-seq data

We applied the method described in “[Reconstructing RNA secondary structures](#)” to each length of cotranscriptional SHAPE-Seq data available with the parameter set found in “[Benchmarking](#)”. Lengths where total mapped read counts are less than 2,000 were not used in R2D2 predictions. For each structure predicted, free energies were calculated using RNAstructure-efn2.

Software implementing this method were run with Python 2.7.12 through Anaconda 2.4.1 (64-bit) and R version 3.2.2. Images and movies were made with ffmpeg version 3.1.3, ImageMagick 7.0.3-0 Q16 x86_64, and iMovie v10.1.2. Version 5.8.1 of RNAstructure was used for the *partition*, *stochastic*, *efn2*, and *ct2dot* methods. RNAbows was used to visualize R2D2 2D predictions ([Aalberts and Jannen, 2013](#)). See “[Data and Software Availability](#)” for location of code used in this study.

Minimum free energy folding pathway prediction

Each length of the SRP RNA sequence was folded with RNAstructure-Fold method without SHAPE restraints to generate the minimum free energy folding pathway. Images of the minimum free energy structures were made into a movie with RNAstructure-draw and ffmpeg. Free energy calculations were done with RNAstructure-efn2. The SHAPE-directed MFE folding pathway prediction was

done similarly, but with ρ reactivities and $m = 1.1$ and $b = -0.3$ (Loughrey et al., 2014) for lengths where SHAPE data was available in specified datasets.

Software implementing this method were run with Python 2.7.12 through Anaconda 2.4.1 (64-bit). Images and movies were made with ffmpeg version 3.1.3, ImageMagick 7.0.3-0 Q16 x86_64, and iMovie v10.1.2. Version 5.8.1 of RNAstructure was used for the *Fold* method to predict MFE structures. See “Data and Software Availability” for location of code used in this study.

KineFold predictions

KineFold cotranscriptional folding pathway predictions were performed using the KineFold executable with ‘co-transcriptional fold’ with a new base added every 20 ms, no pseudoknots, and freely crossing entanglements. KineFold executable was used and can be downloaded from: <http://kinefold.curie.fr/download.html>. For each structure in KineFold’s .rnm output, the free energy was calculated using RNAstructure-efn2. KineFold simulations were also performed with 40 s total simulation time to test if the RNAP roadblocking strategy in cotranscriptional SHAPE-seq (30 s of transcription followed by SHAPE probing) explains differences between R2D2 and KineFold. See “Data and Software Availability” for location of code used to run KineFold and analyze .rnm output.

Using R2D2 predictions to inform all-atom folding pathway simulations

To assess the feasibility of the different hypothetical folding pathways in the full three-dimensional context of the folded RNA, the R2D2 secondary structures were used to restrain all-atom molecular dynamics simulations of each proposed transition pathway. Base-pair constraints for the pre- and post-folding transition were defined using the consensus (base pairs that occur in $\geq 50\%$ of the 100 iterations) R2D2 secondary structures at length 109 and 110 nt respectively. To avoid over-constraining the simulation, only those base-pairs that occurred in over 50% of the reconstructions were enforced with explicit folding restraints. It should be noted that non-restrained bases can still form base-pairs according to the all-atom energy potential. While all pathways start from the same 109-nt folding intermediate (Figures 2F and 5A–5D), each pathway then dictates a unique order in which the base pairing pattern must rearrange to arrive at the final 110-nt native fold (Figure 2G). All-atom simulations employed the GROMACS 2016 software package (Abraham et al., 2015), using the Amber-99 force field (Wang et al., 2000) with Chen-García modifications for RNA bases (Chen and García, 2013), the modifications of Case and co-workers for the backbone phosphate (Steinbrecher et al., 2012), the TIP4P-EW water model (Horn et al., 2004), and the Joung & Cheatham parameters for potassium chloride ions (Joung and Cheatham, 2008).

Simulations employed truncated dodecahedral boxes of ~ 15 nm radius, containing the 110 base RNA, 74,428 TIP4P-EW H_2O ’s, 1,559 K^+ and 1,450 Cl^- ions to mimic 1M excess salt conditions to give a total of 304,265 atoms. Long-range interactions beyond 10 Angstroms were calculated using PME with a grid size of 0.16 nm. A constant pressure of 1 atm was maintained using the Berendsen barostat (Berendsen et al., 1984) with a time constant of 1.0 ps, and a constant temperature of 450K was maintained using the V-rescale thermostat (Bussi et al., 2007) with a time constant of 0.1 ps. The leapfrog Verlet integrator with a 2-fs timestep was used, with the total production length of each simulation being 100-500 ns, leading to a cumulative total of $> 5 \mu\text{s}$ of simulations.

Base-pairs were restrained using a piecewise flat-bottomed harmonic restraint of strength 0.5 kcal/mol between central H-bond donor/acceptor of natively paired bases. This restraint becomes linear at distances greater than 4 Angstroms. The strength and distance dependence of the restraints was chosen to be strong enough to facilitate formation of long-range interactions in ~ 100 ns simulations, but not strong enough to significantly unfold other sections of the RNA in the process. Elevated temperatures were used to increase RNA flexibility and decrease the amount of computational time needed to sample each proposed transition pathway. This arrangement ensured that individual folding attempts would simply stall if two restrained bases could not physically get close enough to form a new basepair in the 3D context of each folding intermediate.

The 110-nt RNA chain was initially equilibrated until all base pairs observed in $> 50\%$ of the stable folding intermediate (109 nt R2D2 2D prediction) were stably formed. At this point, new restraints from the 110-nt natively folded transcript were added 2-3 base-pairs at a time. Each new set of restraints were simulated at least 10 ns until they were successfully formed, at which point the next set of new restraints were added. This cycle was repeated until all bases were successfully paired in the RNA’s native fold. Simulations that had still not achieved any new base-pairs within 5 successive cycles (i.e., 50 ns) after adding restraints were considered “stalled” and not simulated any further. Eight separate attempts were made to simulate each of the 4 proposed pathways and two mutants studied. Each individual folding trajectory therefore ranged from 100-500 ns depending on stalling, and successful folding pathways exhibited at least 6/8 successfully folded trajectories while pathways deemed “unfeasible” always exhibited zero successful attempts.

Four potential folding pathways were simulated. In the “inside-out” pathway, the formation of the extended native helix proceeds by extending the predicted central helix along its axis, unraveling H1 and H3 during this progression, and eliminating the need for forming an initial long-range contact between the RNA ends (Figure 5A). In the “kissing loop” mechanism, it was proposed that complementary, unpaired loop bases within the H1 and H3 hairpins could form an initial long-range “kissing complex,” which could then seed formation of the hybrid helix in a strand rearrangement process (Figure 5B). This hypothesis is attractive because kissing-loop interactions are known to be rapid and stable ways to form long-range RNA interactions in RNA gene regulation and retroviral replication (Kolb et al., 2000; Paillart et al., 2004). A toehold strand exchange mechanism

was also explored, in which the free 3' end of the nascent RNA chain initially hybridizes with unpaired bases in the loop of H1, seeding a sequential unfolding pathway where strands of H1 and H3 are exchanged with each other to rehybridize into the final extended native helix (Figure 5C). Finally, we also explored the “early toehold” mechanism which could initiate at different exposed bases of H1 before H3 is fully formed (Figure 5D).

QUANTIFICATION AND STATISTICAL ANALYSIS

Cotranscriptional SHAPE-seq reactivities were quantified based on a statistical model using Spats v.1.0.1 (<http://luckslab.github.io/spats/>) as described in [reactivity calculation](#) and as previously described (Watters et al., 2016a). The RNase P assay gels were visualized by phosphorimaging on the Typhoon (GE Healthcare), and bands were quantitated using ImageQuant (GE Healthcare) as described in [RNase P assay](#) and [Figure S5G](#).

Additional resources

Detailed protocol

The detailed protocol is provided in [Methods S1](#).