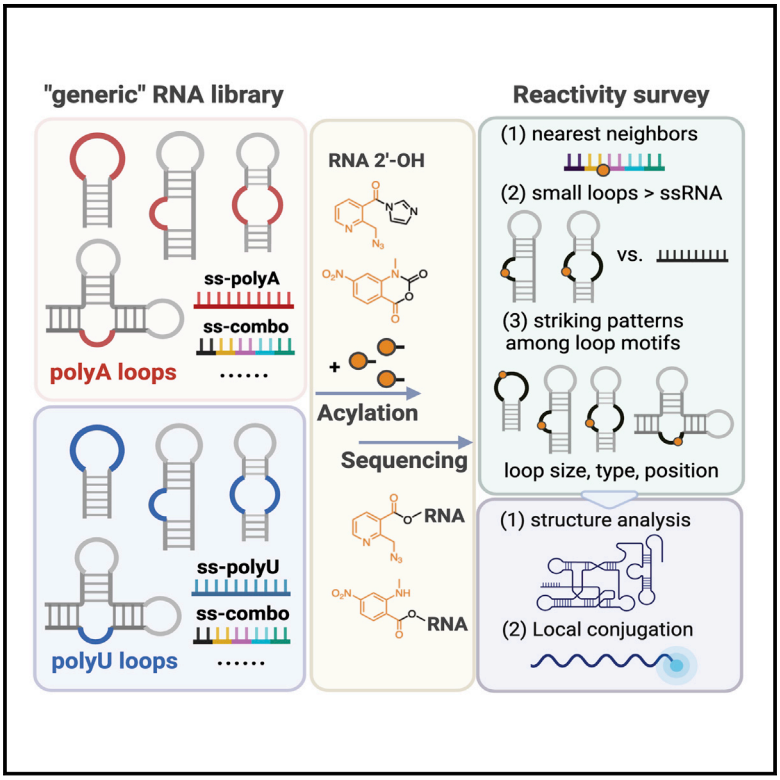
Resource

Cell Chemical Biology

Acylation probing of "generic" RNA libraries reveals critical influence of loop constraints on reactivity

Graphical abstract



Highlights

- Nearest neighbors strongly influence the reactivity of single-stranded RNA sequences
- Small bulge and internal loops are much more reactive than single-stranded RNAs
- Nucleotides in small loops in general react more rapidly than in large loops
- The overall reactivity patterns can guide RNA conjugation and structure mapping

Authors

Lu Xiao, Linglan Fang, Eric T. Kool

Correspondence

kool@stanford.edu

In brief

Xiao et al. surveyed RNA 2'-OH acylation reactivity in "generic" RNA libraries, revealing striking reactivity patterns and critical influence of nearest neighbors and loop constraints. These findings provide insights into RNA designs for high-yield local conjugation and new data to refine RNA structure analysis.

Xiao et al., 2022, Cell Chemical Biology *29*, 1341–1352
August 18, 2022
https://doi.org/10.1016/j.chembiol.2022.05.005

CellPress

Cell Chemical Biology

CellPress

Resource

Acylation probing of "generic" RNA libraries reveals critical influence of loop constraints on reactivity

Lu Xiao,[1] Linglan Fang,[1] and Eric T. Kool[1,2,*]
[1]Department of Chemistry and ChEM-H Institute, Stanford University, Stanford, CA 94305, USA
[2]Lead contact
*Correspondence: kool@stanford.edu
https://doi.org/10.1016/j.chembiol.2022.05.005

SUMMARY

The reactivity of RNA 2'-OH acylation is broadly useful both in probing structure and in preparing conjugates. To date, this reactivity has been analyzed in limited sets of biological RNA sequences, leaving open questions of how reactivity varies inherently without regard to sequence in structured contexts. We constructed and probed "generic" structured RNA libraries using homogeneous loop sequences, employing deep sequencing to carry out a systematic survey of reactivity. We find a wide range of RNA reactivities among single-stranded sequences, with nearest neighbors playing substantial roles. Remarkably, certain small loops are found to be far more reactive on average (up to 4,000-fold) than single-stranded RNAs, due to conformational constraints that enhance reactivity. Among loops, we observe large variations in reactivity based on size, type, and position. The results lend insights into RNA designs for achieving high-efficiency local conjugation and provide new opportunities to refine structure analysis.

INTRODUCTION

The folding of RNA into complex structures is essential for many cellular processes, including protein binding, signaling, splicing, transcription, and translation (Ganser et al., 2019; Sharp, 2009). Accurate analyses of RNA secondary structures are important in deducing structure-function relationships in these processes, and such deep insights into structure are crucial for the study of cell biology. Methods for determining RNA structures include X-ray crystallography (Moore and Steitz, 2003; Robertus et al., 1974), cryoelectron microscopy (cryo-EM) (Gopal et al., 2012), nuclear magnetic resonance (NMR) (Fürtig et al., 2003), and computational folding prediction algorithms (Gruber et al., 2008; Mathews et al., 2004; Reuter and Mathews, 2010; Wu et al., 2015) that make use of experimental information as constraints. In addition to these, chemical probing of RNA is one of the essential tools for providing experimental data of folded structure and, unlike some of the above methods, has the benefit of being able to provide information in the context of biological settings, such as intact cells (Ding et al., 2014; McGinnis et al., 2015; Spitale et al., 2015). Chemical probing typically makes use of reagents that react with RNA in a structure-selective way, thus allowing experimental correlation with folding prediction algorithms (Cordero et al., 2012; Deigan et al., 2009; Hajdin et al., 2013). Important examples of structure-selective chemical probes for RNA include psoralens, which react with pyrimidines in helical contexts when triggered by light (Lu et al., 2016); alkylating agents, such as dimethylsulfate, which preferentially reacts with unpaired adenine and cytosine bases (Tijerina et al., 2007); carbodiimides, which selectively react with unpaired guanine and uridine bases (Wang et al., 2019); and kethoxal reagents, which react with guanine in unpaired contexts (Weng et al., 2020).

One of the most widely adopted chemical probing methods for RNA is that of 2'-OH group acylation (Boerneke et al., 2019; Merino et al., 2005; Velema and Kool, 2020; Wilkinson et al., 2005). An advantage of this approach is that one can obtain information at essentially every position of an RNA strand, due to the ubiquity of 2'-OH groups. This strategy is well established for chemical probing of RNA structures, dynamics, and interactions, following the development by Weeks of the method selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (Merino et al., 2005; Wilkinson et al., 2005). This method takes advantage of an electrophilic acylation reagent, such as 1-methyl-7-nitroisatoic anhydride (1M7) or azidomethyl-nicotinoyl-acylimidazole (NAI-N_3), to selectively react with 2'-OH groups at unpaired or conformationally flexible nucleotides, forming 2'-O-ester adducts. The adducts can be read out by cDNA truncations or mutations that occur during reverse transcription (RT). The truncations or mutations can be analyzed by electrophoresis or deep sequencing to identify the reactive positions, generating information at nucleotide resolution along the strand (Busan et al., 2019). Moreover, such probing methodologies can also be used to identify RNA noncovalent interactions, such as protein binding, by comparison of reactivity profiles of RNA alone to that in cells (Spitale et al., 2015; Weidmann et al., 2021). Thus, the

acylation-based probing of RNA has been adopted as a powerful analytical tool, and a deeper understanding of this reactivity may contribute to improved tools and methods.

In addition to probing RNA structure, RNA acylation has more recently been developed as a tool for RNA conjugation. Classical mapping reagents give low yields of reaction with RNAs, which is useful for structure analysis but is not practical for high-yield conjugation reactions. This has led to the development of new classes of reagents that achieve high chemical yields for RNA acylation, such as acylimidazoles (Habibian et al., 2019; Kadina et al., 2018; Knorre et al., 1965; Park et al., 2019; Spitale et al., 2013; Velema et al., 2018) and alkylamino-substituted isatoic anhydrides (Fessler et al., 2021; Ursuegui et al., 2014). These developments have enabled ready installation of groups into RNA, such as biotin for affinity purification and fluorophores for quantitation and imaging. Also under development are methods for site-localized acylation (Xiao et al., 2020, 2021). While in some applications, stochastic reaction along an RNA strand has proven useful for such conjugation, it is also desirable in some instances to be able to direct acylation to specific sites, thus providing better molecular homogeneity and less interference with RNA function. Such RNA conjugation strategies could benefit from a deeper understanding of structure-based 2'-OH acylation reactivity by potentially identifying highly reactive structural motifs to improve yields and selectivity.

It has long been observed that RNA 2'-OH acylation reactivity is not simply binary with the "unpaired" or "paired" nature of a given nucleotide but rather varies among different unpaired motifs (McGinnis et al., 2012). Researchers aiming to improve RNA structure analysis have utilized previously measured quantitative SHAPE data from biological RNAs to derive improved algorithms for folding analysis (Cao and Xue, 2021). This can lead to better success in predicting accurate structures and analyzing tertiary interactions, such as occurs in pseudoknots. Such studies to date have relied on a limited set of data, covering a relatively small fraction of possible loop sequences. Since it is known that common-sized loops (such as tetraloops) can adopt sequence-dependent cross-ring interactions (Thapar et al., 2014), this suggests that the use of a limited but sequence-varied dataset will obscure some of the more generic positional effects on reactivity. For these reasons, we suggest that modified mapping algorithm development could also benefit from systematic data that can serve as a baseline, obtaining reactive information about position-dependent reactivity with homogeneous loops in uniform contexts. In addition, previous data have been measured primarily with one reagent, leaving open the question of how different reagent classes might react differently in variable structure contexts. These facts suggested to us that new experimental data on baseline information regarding loop structural contexts could be useful in multiple respects, both for structural analysis of biologically active RNAs and for future design strategies for RNA conjugation.

Overall, there remains a gap in systematic information about variations of positional acylation reactivity across different secondary RNA structures. This prompted a number of questions about these differences. For example, do different types of loops—omitting differences in sequence—present distinct levels of reactivity to acylation? We imagined that structural factors that modulate the accessibility and the nucleophilicity of the 2'-OH groups could differ between varied loop motifs. Similarly, do all 2'-OH groups within a loop have the same inherent reactivity, or does it vary along the loop? Does this depend on loop size or pyrimidine and purine composition? And how does acylation reactivity compare between single-stranded conformationally flexible RNAs and loops, which are more constrained? Finally, to what degree do base stacking and nearest-neighbor interactions play a role in reactivity? To date, the details of some of these questions remain poorly understood.

In this work, we address these issues by carrying out a systematic survey of acylation reactivity in two designed homogeneous loop-RNA libraries via acylation methods integrated with RT stops and deep sequencing. Two common RNA-acylating reagents, 1M7 and NAI-N_3, were investigated with these libraries containing 43 RNAs with systematically varied hairpins, bulge loops, internal loops, and junctions of varied size as well as single-stranded sequences. We find that the reactivity varies along single-stranded RNAs with essential effects of base identity and nearest neighbors. We also find systematic positional variations in reactivity among different loop types and sizes, serving as distinct fingerprints of loop types. Significantly, we find that certain loop types are much more reactive than completely unconstrained RNAs, arguing for the importance of accessibility and loop conformational constraints in determining reactivity. We further apply the reactivity observations to design a prototype for a highly reactive RNA "acylation tag." Our work provides baseline data to refine the analysis of RNA structure and interaction mapping by acylation and provides insights into future RNA designs for achieving high levels of local conjugation.

RESULTS

Strategy for a quantitative survey of 2'-OH acylation reactivity across RNA structural motifs

To assess the quantitative origins and variations of acylation reactivity, we employed deep sequencing to conduct a systematic survey of reactivity across two RNA libraries having loops of uniform sequence (A-loops and U-loops; Figure 1). Each RNA library was constructed to contain 22 systematically varied RNA structure elements, including hairpins (2- to 7-nt loops), bulge loops (1- to 5-nt loops), internal loops (1 × 1, 2 × 1, 2 × 2, 3 × 2, and 3 × 3 loops), and three-way junctions (with 1- to 3-nt loops). Closing base pairs (adjacent to loops) were C-G/G-C to stabilize folded structure. The structured RNAs in A-loop and U-loop libraries shared identical stem sequences but incorporated homogeneous loops with polyadenosine ("A-case") and polyuridine ("U-case") composition. We chose uniform loop sequences to remove stacking and steric variability along the loop positions, to allow us to analyze positional effects without sequence variation. A and U loops were employed because of their divergence in stacking proficiency and size; A, a two-ring purine, is the strongest stacking base of the four canonical nucleobases, while U, a single-ring pyrimidine, is the weakest (Turner and Mathews, 2010). Also worth noting is that A is the most common nucleotide in predicted tetraloops in biological RNAs, while U is the most frequent nucleotide in larger loops (Danaee et al., 2018). Each RNA in the library also contained a unique 3-nt barcode in the stem near the end to allow it to be distinguished among the set via our bioinformatics

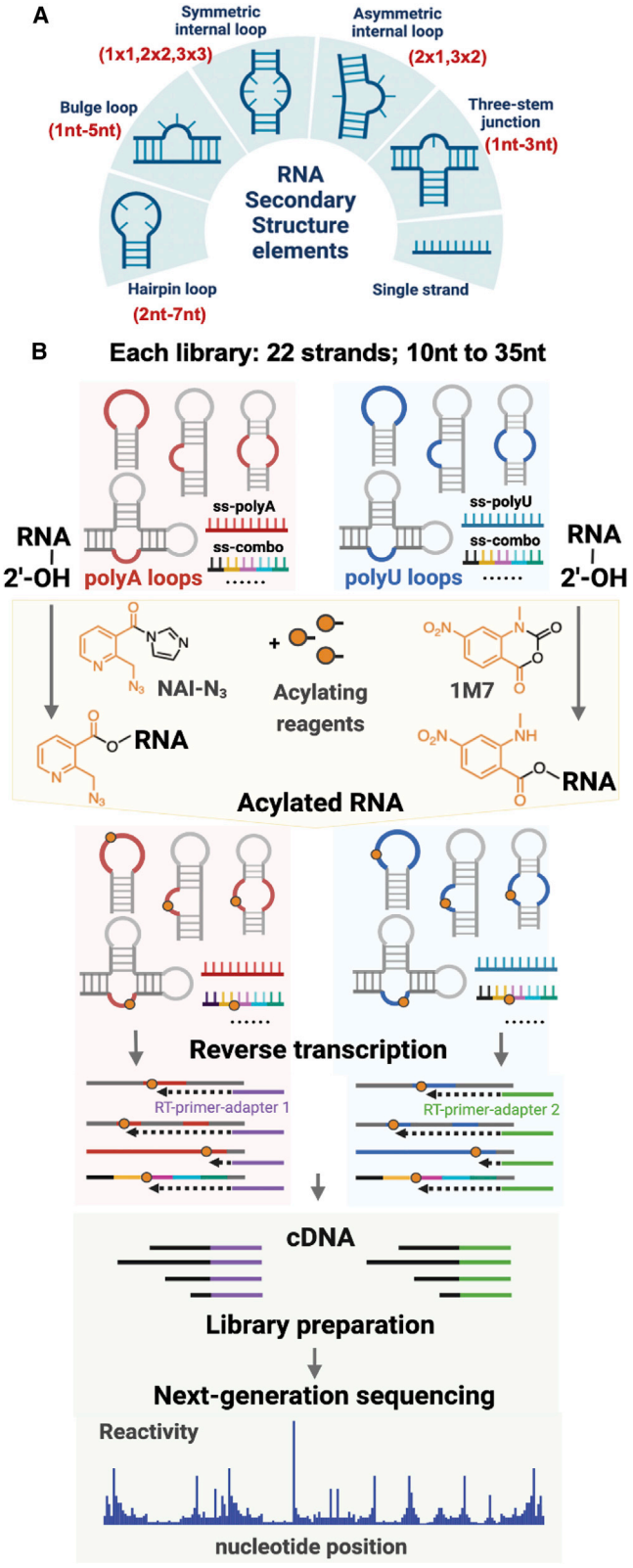


Figure 1. Strategy for a quantitative survey of acylation reactivity across RNA structural motifs

(A) RNA secondary structure elements studied in this work.

(B) Schematic of the strategy for probing the libraries with acylation and sequencing. Two RNA base libraries having uniform loop sequences (A-case and U-case) were each constructed with 22 RNA elements with varied loop sizes, including hairpins, bulge loops, internal loops, and three-way junctions, as well as single-stranded RNAs. RNAs of the same type and size in each library shared an identical stem sequence containing a barcode as a unique identity code. Loops were composed of polyA loops in the A-case set and polyU sequences in the U-case set, respectively, to avoid sequence-dependent variation in reactivity. Two common acylating reagents, 1M7 and NAI-N_3, were reacted with RNA libraries at 37°C in folding buffer, followed by reverse transcription, library preparation, and next-generation sequencing analysis to quantify the positional reactivity information.

pipeline (Figure S3H). Finally, three fully single-stranded RNAs (ssRNAs) were also included in the analysis to probe fully unconstrained and flexible sequences as a baseline. One ssRNA contains polyA sequence, one polyU, and a third contains all possible nearest-neighbor steps. All the sequences were interrogated together for consistent exposure to reagents (keeping A-loop and U-loop series separate to avoid loop-loop interactions) under conditions that stabilized individual folding. RNAs were folded separately before being combined immediately prior to mapping.

All RNA sequences are given in the supplemental information file; they were designed using the MFOLD and RNAFOLD (Gruber et al., 2008; Zuker, 2003) folding algorithms. The RNAs were synthesized chemically and individually purified by gel electrophoresis. One representative folded structure for the sequence context was tested by SHAPE mapping with electrophoretic analysis, confirming the predicted fold, and was utilized for reaction condition optimization (Figure S1A). Probing experiments were carried out with two common types of acylating reagents, 1M7 and NAI-N_3. These represent the two standard structural classes of acylation reagents and also provide observation of the effect of varied reactivity; 1M7 has a half-life in water of 14 s (Mortimer and Weeks, 2007), while that of NAI-N_3 is ~30 min (Lee et al., 2017; Spitale et al., 2013). Both reagents were reacted with RNAs under the optimized condition of 37°C for 10 min in a pH 7.5 folding buffer containing 100 mM HEPES, 100 mM NaCl, and 6 mM $MgCl_2$ (Figure S1A). Different concentrations of reagents were tested (Figure S1A) to identify optimal mapping conditions. We also conducted mock (DMSO) treatment as a control for both RNA libraries to account for native stops generated by RT in the context of these sequences. For the acylation analysis, we utilized specific RT primers containing distinct adapters for each reaction condition, followed by library preparation and deep sequencing with all combined reactions to survey the acylation across all RNA strands. Sequencing data and the positional reactivity of RNAs were analyzed through a bioinformatics pipeline adapted and modified from previous work (Flynn et al., 2016; see supplemental information).

Reactivity survey of single-strand RNAs

To measure the intrinsic reactivity of nucleotides and nearest neighbors, we first investigated the positional reactivity of a designed ssRNA containing all 16 dinucleotide steps (ss-combo) (Figure 2A). Few prior studies exist that analyze acylation reactivity of purely ssRNAs (Wilkinson et al., 2009). We evaluated the reactivity of this ssRNA with NAI-N_3 and 1M7; this provided

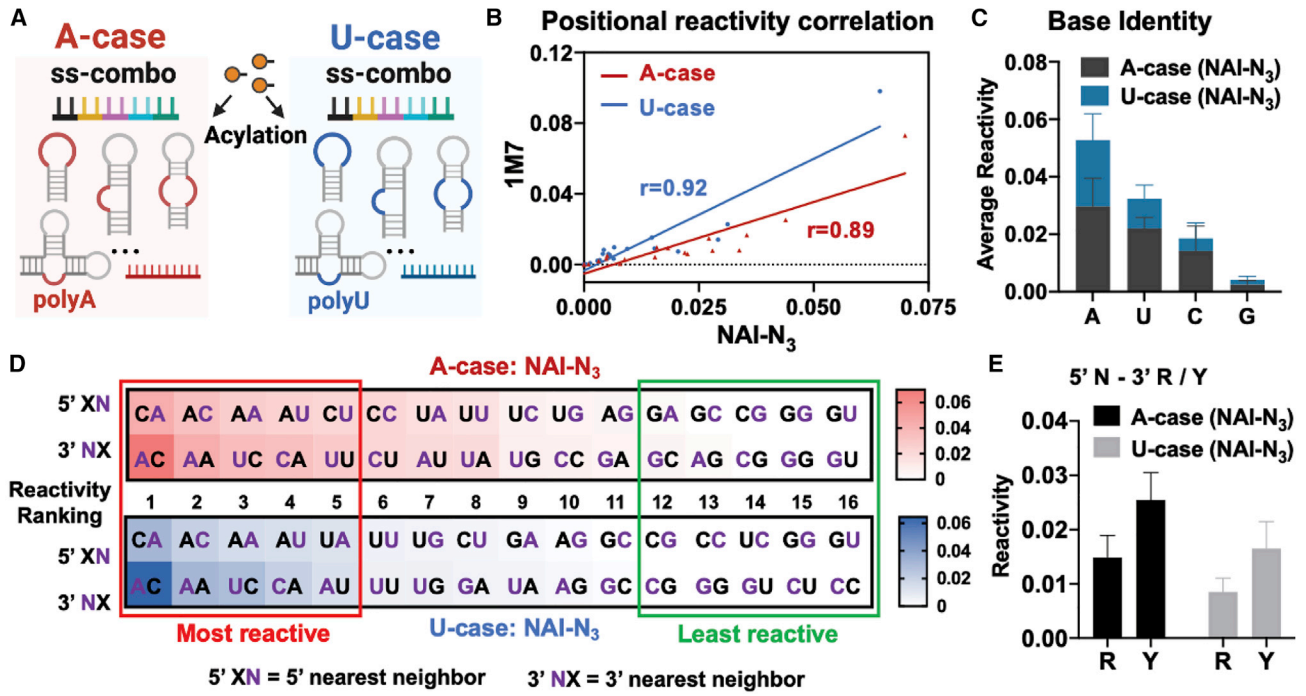


Figure 2. Reactivity survey of single-stranded RNAs

(A) Schematic of reactivity survey of single-stranded RNA containing all 16-nt steps (ss-combo) in both A-case and U-case RNA library.

(B) Positional reactivity correlation of 1M7 and NAI-N_3 along the ss-combo RNA sequence. "r" represents Pearson correlation coefficient.

(C) Base identity effect on the reactivity in the ss-combo RNA.

(D) Reactivity ranking of nearest neighbors along ss-combo RNA sequence. N, acylated nucleotide; X, 5' or 3' nucleotide to the acylated one.

(E) The steric effect of 3' nucleotide on reactivity in the ss-combo RNA. The reactivity of the 5' nucleotide shown was affected by its 3' nucleotide either in the presence of purine (R) or pyrimidine (Y).

Data are represented as mean ± SEM.

well-correlated results for the two reagents with a high Pearson correlation (r = 0.89 in A-case set; r = 0.92 in U-case) (Figure 2B). The results showed a clear reactivity trend with base identity, A > U > C > G (Figure 2C), which is similar to the nucleotide reactivity reported for NAI-N_3 in 16,000 polyA$^+$ RNAs in mouse stem cells (A > U > G > C) (Spitale et al., 2015). The ss-combo RNA data also revealed widely different reactivity among the nearest neighbors (Figures 2D, S2A, and S2B). The top four most reactive nucleotide steps in common are CA, AC, AA, and AU, while the four least reactive ones are GG, GC, CG, and GU (Figures 2D, S2A, and S2B). Note that this does not correlate well with reported effects of nucleobase identity, as early studies observed lower acylation levels of C relative to the other nucleotides in biological RNAs (Wilkinson et al., 2009). We compared the measured stacking propensities of the nearest neighbors (Brown et al., 2015) with their reactivities, but this showed no apparent correlation (Figures S2C–S2E), revealing the existence of effects beyond stacking alone, which may not be surprising, given the fact that the ssRNA is very likely above its helical (stacked) transition temperature. Next, we compared the average acylation reactivity of each nucleotide in the presence of either purine or pyrimidine with the 3' side of its reactive hydroxyl. Nucleotides with a pyrimidine at the 3' position showed relatively higher reactivity than with a purine (average 1.8-fold difference), which may be due to the steric effect of larger purines blocking access to the nearby 2'-OH group (Figure 2E).

Reactivity comparison of single-stranded RNA and loop RNA

To first confirm the expected folding statuses of structured RNAs in the library, we compared the average reactivity of stems (duplex RNAs) and loops of the designed RNAs in each set. Duplex RNAs are known to have low SHAPE reactivity in well-folded structures. We found that the reactivity of nucleotides in predicted loops are substantially higher than those in stems for the different RNA motifs. Loop nucleotides reacted at 4- to 9-fold higher levels on average than stems for both acylating reagents, indicating that the folded RNA statuses are predominant in each RNA library (Figure S1B). Inspection of individual library members revealed similar stem reactivity of each strand (Figure S1C). The stem/loop reactivity ratio (R_{stem}/R_{loop}) per strand is consistent with the average loop/stem ratio analysis for all but five hairpin cases containing larger loops (HP-5A/U, HP-6A/U, and HP-7U), where loop reactivity is quite low from our SHAPE sequencing data (Figures S1D and S1E). To further confirm the folding of these hairpins, we conducted independent dimethylsulfate (DMS) probing for A-loop hairpins in buffer (Figure S1F) and comparative SHAPE probing for U-loop hairpins in buffer (folding) and water (denaturing) conditions (Figures S1G and S1H). All hairpin RNAs under folding condition showed reactivity maps consistent with the predicted stems and loops. Thus, the probing data together confirm the expected folds for all library members.

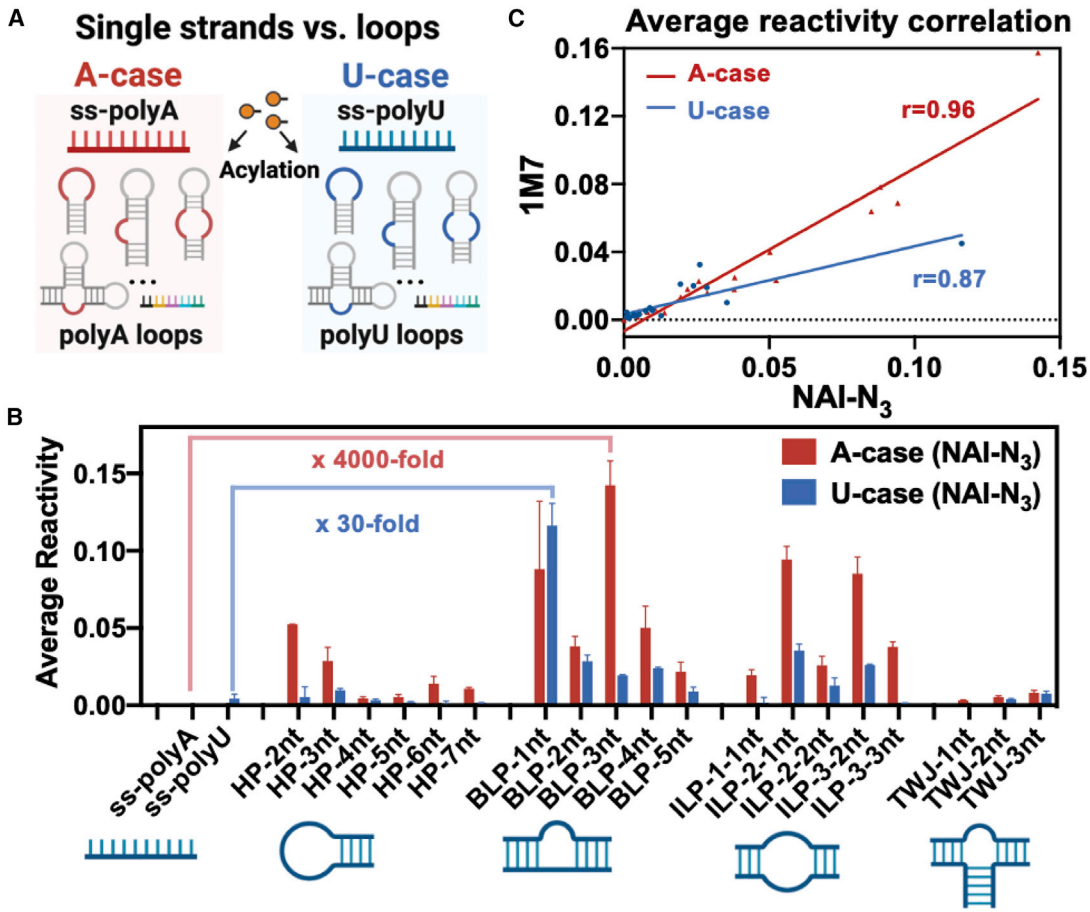


Figure 3. Reactivity comparison of single-stranded RNA and loop RNA

(A) Schematic of reactivity comparison of single-stranded RNAs (ss-polyA and ss-polyU) versus RNA loops.

(B) Bar graph of average reactivity of loop residues in different types of loops (polyA loops and polyU loops) as well as single-stranded RNAs (ss-polyA and ss-polyU).

(C) Average reactivity correlation of two acylating reagents, 1M7 and NAI-N_3, in all types of RNA loops and ssRNAs in these libraries.

Data are represented as mean ± SEM.

To further assess the degree to which RNA flexibility and loop constraints affect reactivity, we compared average acylation reactivities of purely ssRNAs (ss-polyA and ss-polyU) and A/U loops in the structured RNA sets (Figure 3A). We found that the average reactivity in most loops far exceeded that in single-stranded (fully flexible) sequences, suggesting that loop structural constraints on average expose specific 2'-OH groups for greater accessibility to acylation (Figures 3B and S4A). Remarkably, the average reactivity of the most reactive loop in the A-case (the 3-nt A-bulge loop) on a per-nucleotide basis was over 4,000-fold higher than ss-polyA. For U-loop cases, the most reactive loop is also a bulge loop (the single-nucleotide bulge), and it reacts at over 30-fold-higher levels than ss-polyU. Also interestingly, the average reactivity of ss-polyU is higher than ss-polyA, likely due to the larger size of adenine and lesser stacking of ss-polyU, which is expected to exist in a dynamic random coil conformation at the reaction temperature of 37°C (Figures 3B and S3A). Literature reports have established T_m values of polyA and polyU of ~41°C and ~5.8°C (Lipsett, 1960; Richards, 1968), respectively, indicating that the ss-polyA sequence is likely partly helical, which is expected to confer significant steric protection (see discussion).

Next, we surveyed reactivity differences among distinct loop types and found large and surprising variations. Reactivity patterns for the two reagents were similar, with a high correlation of average reactivity with each type of RNA (r = 0.96 for A-case series; r = 0.87 for U-case; Figure 3C). Bulge loop and internal loops were consistently found to be more reactive than the hairpin loops and three-way junction loops, which further encouraged us to investigate the reactivity features among loop types (Figure S3B).

Reactivity survey across loops in A-case and U-case RNA libraries

To survey 2'-OH acylation reactivity in the context of different RNA secondary structures, we analyzed the detailed reactivity of the two reagents, NAI-N_3 and 1M7, among different types of loops (hairpin loop, bulge loop, internal loop, and three-way junction) with systematically varied polyA/polyU loop sizes (ranging from 1 nt to 7 nt; Figure 4A). The two reagents provided

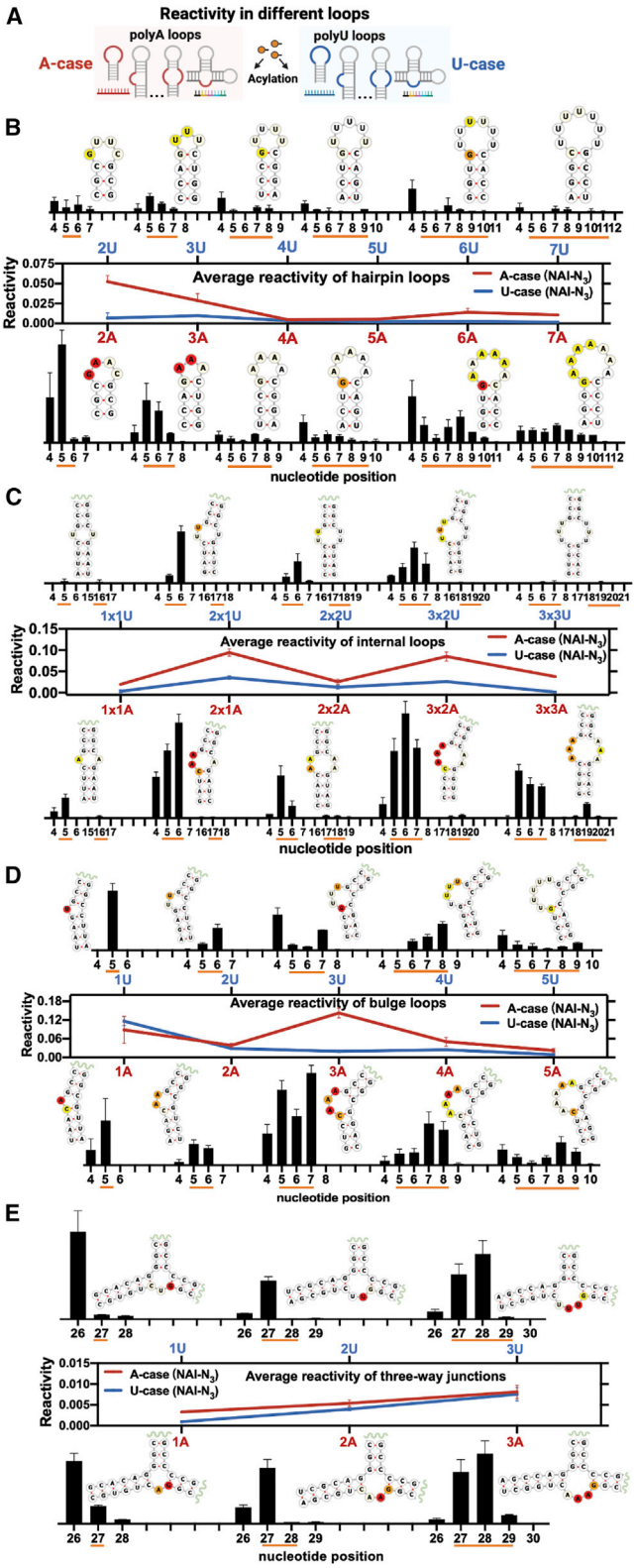


Figure 4. Characteristic reactivity patterns of NAI-N_3 for hairpin, internal, bulge, and three-way junction loop motifs, including closing bases

(A) Schematic of reactivity survey across different types of RNA loops.

(B) Average and positional acylation reactivity in hairpin loops (2–7 nt) and closing bases.

(C) Average and positional reactivity in internal loops (1 × 1, 2 × 1, 2 × 2, 3 × 2, and 3 × 3) and closing bases.

(D) Average and positional reactivity in bulge loops (1–5 nt) and closing bases.

(E) Average and positional reactivity in three-way junctions (1–3 nt) and closing base pairs. The numbers underlined show the nucleotides in the loop, and the adjacent ones are the closing bases. The qualitative reactivity in the structure is shown in order of color intensity from most reactive to least reactive (red > orange > yellow > light yellow).

Data are represented as mean ± SEM. Note that data shown were obtained with reagent NAI-N_3; see Figure S4 for data with 1M7.

quantitatively nearly identical results in the A-case RNA library, with a Pearson correlation (r) value of 0.97 for average loop reactivity and 0.86 for positional reactivity of all loop nucleotides plus closing bases (Figures S3C–S3E). In the U-case set, the two reagents were correlated in average loop reactivity (r = 0.86); however, they showed relatively low correlation for specific positional reactivity (r = 0.32), due chiefly to higher 1M7 reactivity at three specific sites (Figures S3C–S3F). For simplicity, we display full NAI-N_3 data here, and the 1M7 data are provided in the supplemental information file.

We next conducted a comparative study of acylation reactivity of the loops and the closing base pairs. Overall (with a few exceptions for U loops as noted above), the reactivity patterns for each type of loop are very similar for both NAI-N_3 and 1M7. In the hairpin context, we examined the polyA and polyU loops with sizes ranging from 2 nt to 7 nt (Figures 4B and S4B) (note that 2-nt hairpin loops may not exist as such depending on sequence [Jucker and Pardi, 1995] and could exist as 4-nt loops with the first and last nucleotides unpaired). We noted a strong loop size effect, with small loops being the most reactive. We hypothesize that bond constraints in small loops likely cause the adoption of noncanonical bond rotamers (i.e., widely altered from A-form torsional angles) favoring 2'-OH group exposure. It is also possible that the constraints affect local sugar conformation as well (McGinnis et al., 2012). The effect of nucleotide position in loops is also strong in some cases, with the 5' end of loops showing high reactivity and the 3' end showing less reactivity (in some cases near zero). For example, trinucleotide loops show high reactivity at the first two positions and little at the third. We hypothesize that the rigidity of closing base pairs may directly reduce the freedom of 3' phosphate in the 3'-end loop nucleotide, which could subsequently block its adjacent 2'-OH group from acylation, while the rigidity effect of 5' closing base is alleviated by the sugar 5'-carbon bridge. Examination of the published solution structure of a tri-loop hairpin (Kim and Tinoco, 2001) (PDB: 1I4B; Figure S5A) shows that the loop is more flexible at the 5' end than the 3' end according to b-factor analysis, pointing to greater dynamic accessibility of 2'-OH group at the 5' part of the loop.

Interestingly, the closing base pair of the loops also exhibited a striking asymmetry in acylation, with high reactivity of the paired 5' nucleotide (immediately prior to the loop) over the 3' closing nucleotide (Figures 4B and S3G). Indeed, the 5' nucleotide of the closing pair was in several instances more reactive than any of the loop nucleotides in a given hairpin (Figure 4B). We hypothesize that the asymmetric reactivity difference (5' versus 3' closing nucleotide) may be due to the combination of steric

and flexibility effects. Based on the aforementioned NMR tri-loop RNA structure (Figure S5A), we see that the 2'-OH group of 5' end closing base faces the loop, which is more flexible as shown by the b-factor analysis, while the 2'-OH at the 3' closing nucleotide is directed into the rigid helix region, suggesting more steric hindrance and less accessibility for 2'-OH acylation (Figure S5A). We also examined the NMR structures of additional hairpin loops (Butcher et al., 1999; Jucker and Pardi, 1995; Nomura et al., 2006) and found that the phenomenon is prevalent in other RNA structures as well (Figure S5B).

We further investigated the RNA acylation reactivity in polyA/polyU internal loops with varied loop sizes (1 × 1, 2 × 1, 2 × 2, 3 × 2, and 3 × 3). Internal loops varied surprisingly widely in their reactivity, depending on whether they are symmetric or asymmetric (i.e., same or different numbers of nucleotides on each side). A particularly striking characteristic was significantly higher reactivity in asymmetric internal loops compared with symmetric loops (Figures 4C and S4C). Moreover, reactions occur primarily at nucleotides on the over-numbered side of the loop. This is consistent with data in a recent meta-analysis (Cao and Xue, 2021), and literature reports have shown that asymmetric loops are less stable than symmetric loops due to the backbone distortion with unfavorable free energy for folding (Peritz et al., 1991). We hypothesize that the destabilization tendency in asymmetric loops might favor the accessibility of 2'-OH groups of nucleotides to the acylating reagents. Examination of NMR structures of internal loops (Butcher et al., 1997; Collier et al., 2002; Desjardins et al., 2011; Lescrinier et al., 2003) in the PDB database shows that mismatched bases in symmetric internal loops generally were located inside the helix; in contrast, the extra nucleotide in asymmetric loops led to a distortion, flipping out the bases and therefore likely increasing the exposure of 2'-OH groups for acylation (Figure S5C).

Bulge loops with polyA and polyU composition were tested over the size range of 1–5 nt. The results showed reactivity patterns similar to hairpin loops, with small loops being more reactive but displaying extraordinarily high reactivity for an A_3 loop in the A series (Figures 4D and S4D). The positional effect in bulge loops is minimal, but the asymmetric reactivity in the closing base pair is still seen, likely due to similar factors as in hairpin loops (Figure S5D; Diener and Moore, 1998; Greenbaum et al., 1996; Popenda et al., 2008; Smith and Nikonowicz, 1998). The reactivity of three-way junctions increased with the size of loops (1–3 nt) for both reagents and both RNA libraries (Figures 4E and S4E). Finally, the 5'-favoring effect of nucleotide position was also observed in junction loops, with 5' end reactivity higher than 3' end, and with similar asymmetric closing base pair reactivity patterns.

Surveying the reactivity data as a whole, we can also rank the relative reactivities of these unpaired motifs. The most reactive loops tested here are trinucleotide bulge loops and asymmetric internal loops (Figure 3B). Surprisingly, those composed of adenosine nucleotides were considerably more reactive than uridine nucleotides, despite the smaller size and lower stacking ability of the latter. Some of this difference may be attributed to the inherently higher reactivity of the 2'-OH of adenosine nucleotides over uridine nucleotides (Jash and Kool, 2022). The least reactive loops were those associated with three-way junctions, which may be a result of steric crowding in the converging helices (Figure S5E; Bonneau and Legault, 2014). Finally, and unexpectedly, the fully ssRNAs displayed among the lowest reactivities of all unpaired nucleotides in the set, which strongly suggests that flexibility alone is a poor predictor of reactivity and that loop bond constraints play major roles in determining 2'-OH accessibility and reactivity.

To further assess the generality of our findings, we conducted a meta-analysis of average positional reactivity of different types of loops in biological RNAs using the SHAPE datasets published by Cao and Xue (2021). We observed similar trends of asymmetric reactivity at the over-numbered sides of internal loops, as well as consistently biased reactivity near 5' end of hairpin loops (Figure S6). Thus, the current results are generally consistent with previous data from biological RNAs.

Applications of reactivity data

The above data have established multiple surprising observations about acylation reactivity in different structural contexts. Two of the most striking findings include the discovery that certain small loops are highly reactive in comparison to larger loops and relative to single-stranded RNA as well. Moreover, the data show that purely single-stranded RNA varies in its reactivity with different nearest-neighbor environments. The results suggest possible ways to use this information in designs for improved RNA conjugation (see discussion).

To test these findings again in a different sequence context, we prepared and studied three new RNA constructs: one a purely ssRNA and the other two folded RNAs. For the single-stranded case, we designed a 16-nt sequence with nearest neighbors rearranged to place more reactive positions in the 5' half and less reactive in the other. The ssRNA sequence was situated immediately downstream of a primer binding site for probing reactivity. Acylation reactions with NAI-N_3 were again carried out under standard conditions and analyzed by gel electrophoresis after RT (Figure 5A). The results showed selective acylation at 5' end sites with the predicted reactive nearest neighbors (red part); in contrast, moderate bands occurred at nucleotides with medium reactivity (orange part), and little or no acylation was observed in the least reactive nearest neighbors (green part) (Figure 5B). The selectivity increased with decreasing NAI-N_3 concentration (Figure 5B, last lane), clearly showing this nearest-neighbor-dependent reactivity along the ssRNA.

In a second test, we designed two new RNA constructs with the potential to confer highly active, localized acylation reactivity for site-selective labeling. The constructs contain a bulge loop (the highly reactive A_3 loop and the single-nucleotide-selective U_1 loop identified above) connected to a single-stranded primer binding site for analysis by capillary electrophoresis (CE) (Figure 6A). This CE handle was protected by a complementary DNA during reaction to ensure efficient detection of acylation reactivity in the loops. Also included was a hairpin loop, necessary to include the structural context for the bulge loop. For this hairpin loop, we chose the most unreactive case in our library, namely, the U5 loop. The motifs are compact (22 and 20 nt) and allow direct tests in a new overall context (bulge loop tag [BLT]-A_3/BLT-U_1) of the reactivity contrasts between low-reactive loops and highly reactive cases.

The results showed that both small bulge loops (A_3/U_1) displayed higher reactivity than the larger and presumably more

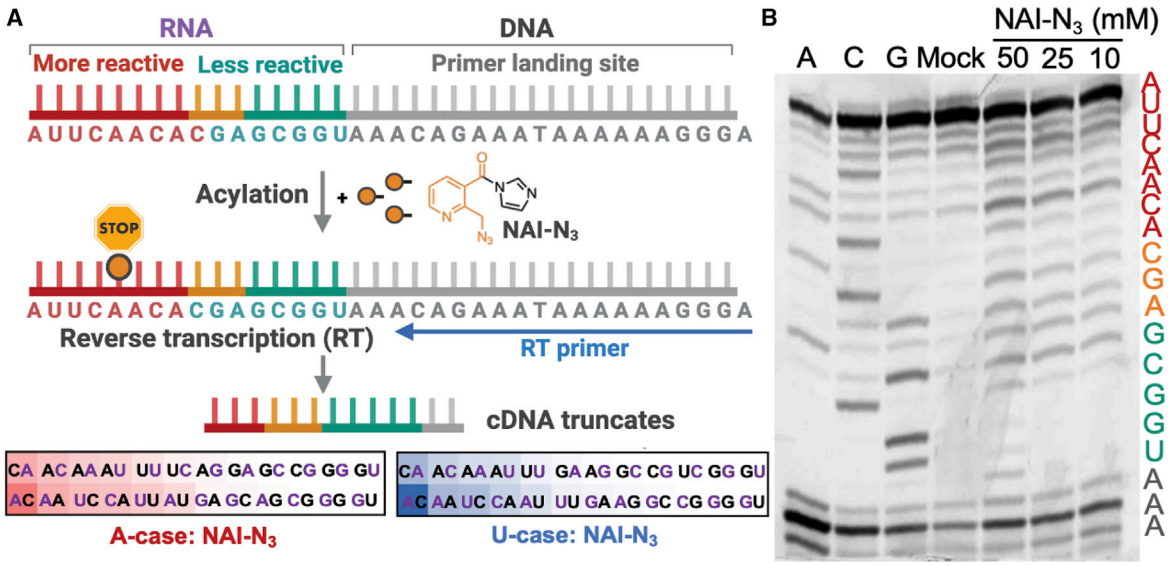


Figure 5. Independent test of single-stranded RNA (ssRNA) varying in its reactivity with different nearest-neighbor environments

(A) Schematic of acylation of a designed ssRNA with more reactive nearest neighbors arranged in the 5' portion.

(B) Gel analysis after RT of the designed ssRNA (sequence at right) reacted with NAI-N_3. Note that RT stops (and corresponding bands) occur at the nucleotide immediately 3' to the reacted position.

flexible hairpin loop (U_5) in both RNA structures (BLT-A_3 and BLT-U_1), consistent with results from our library sequencing data (Figures 6B and 6C). The average reactivity of the A_3 bulge loop was ~3-fold higher than the U_5 hairpin in BLT-A_3, and the U_1 bulge loop reacted ~4-fold more than the U_5 hairpin in BLT-U_1. Minimal reactivity was observed in stem regions of the RNAs. These results further confirm the above findings and reveal the potential of using small bulge loops as RNA tags for site-selective labeling and conjugation.

DISCUSSION

Our data have shown sequence-generic positional effects on RNA acylation reactivity with identified features among specific RNA structural motifs. We are aware of no prior studies of positional reactivities of loops all bearing the same base composition. We find that the general conclusions are largely the same with two reagents of widely differing reactivity. One significant observation here is that the reactivity of (non-loop) ssRNA varies with base identity and depends on whether a purine or pyrimidine is positioned 3' to the sugar undergoing reaction (Figure 2). We hypothesize that the steric effect of larger purines may hinder reagent access to the 2'-OH group nearby. Comparing the purely ssRNAs (ss-polyA and ss-polyU) with the A/U loops in structured motifs, we find that loops (especially small ones) are more reactive than ssRNA, likely due to more frequent and unhindered exposures of 2'-OH groups caused by loop structural constraints (Figure 3).

Comparing the reactivity among loops, the size effect is generally strong. In hairpins and bulges, small loops display higher reactivity on average (Figures 3B, 4B, and 4D), which may be due to bond constraints in small loops leading to non-A-form bond torsion angles and exposing the 2′-OH group (either statically or dynamically) for productive collision with acylating molecules. In internal loops, asymmetric loops exhibit significantly higher reactivity than symmetric loops. Acylation primarily occurs at the over-numbered side, which is consistent with the previous analyses of 5 × 6 and 3 × 6 internal loops with published SHAPE data in the context of biological RNAs (Cao and Xue, 2021; Figure S6B). Structural and biophysical studies of asymmetric loops report them to be less stable than symmetric loops due to backbone distortion (Peritz et al., 1991), which might also facilitate bases flipping out on the over-numbered side, increasing the accessibility of the associated 2'-OH groups (Figures 3B and 4C).

Comparing the nucleotide reactivity within each loop motif, our data have shown two striking positional effects (Figures 4, S3G, and S4B–S4E). First, loop nucleotides near the 5' end are generally more reactive than 3' end in hairpins and junctions. Previous meta-analyses of existing data from biological RNAs have shown a similar reactivity trend in hairpins (Figure S6A; Cao and Xue, 2021). Second, we find that loop-closing base pair nucleotides have high asymmetry in their reactivity, with much higher reactivity of the 5' closing nucleotide over the 3' pairing partner. Examination of published structures suggests that the 5' closing nucleotide has its 2'-OH directed upward into the relatively open and flexible loop, while the partner base has its 2'-OH aiming downward into the helix, with a similarly protected orientation as the rest of the A-form base pairs (Figure S5). NMR structure data also show that loops are more flexible at 5' ends than 3' ends according to b-factor analysis (Kim and Tinoco, 2001), suggesting greater dynamic accessibility of 2'-OH groups on the 5' side.

Flexibility of a given nucleotide has long been associated as a concept with acylation reactivity in RNAs (McGinnis et al., 2012; Wilkinson et al., 2005), as loops react more readily than duplex RNAs and are clearly more flexible as well. In addition, hyperreactive nucleotides have also been noted in studies of biological RNAs, suggesting contributions from localized structural effects,

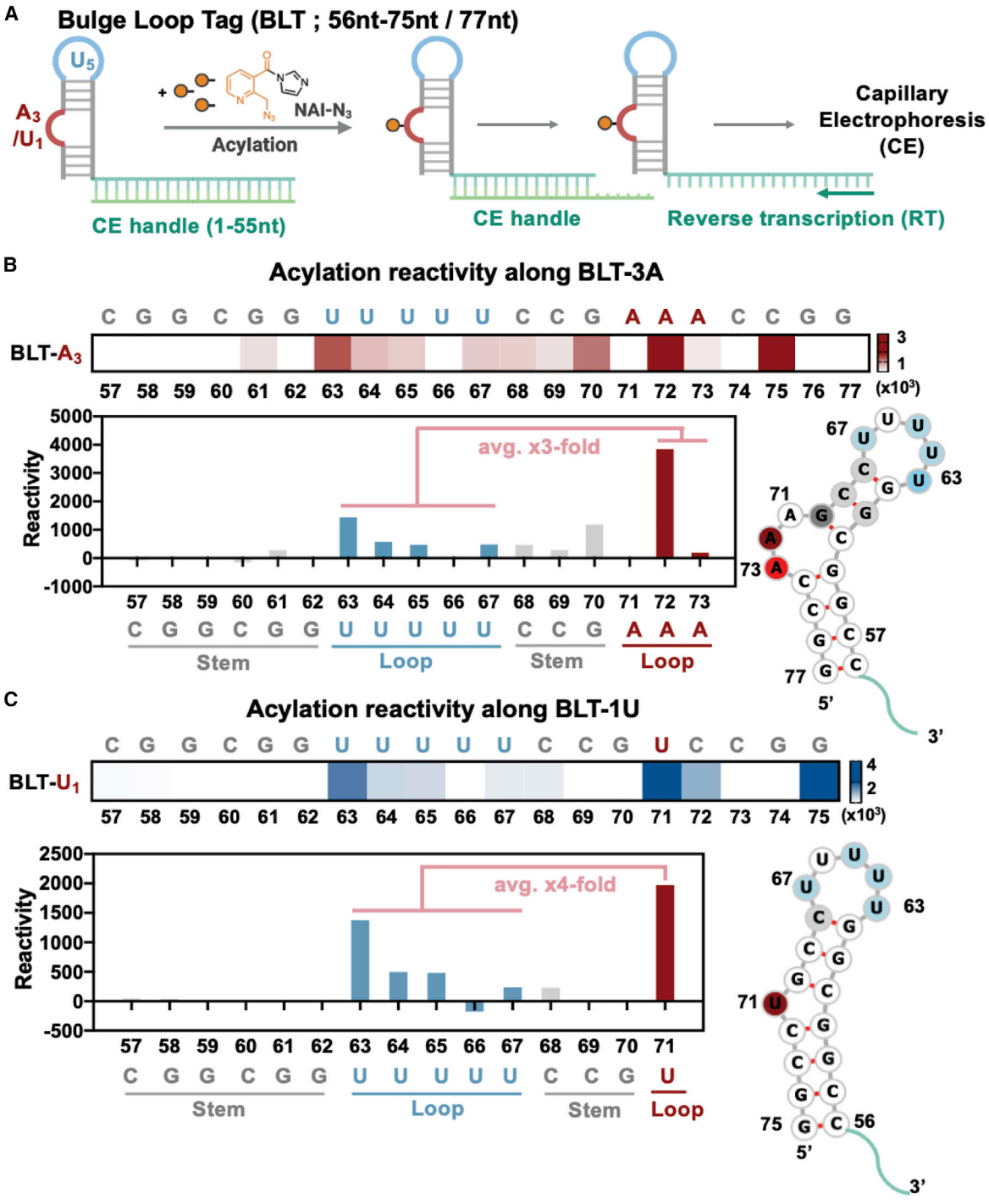


Figure 6. Independent localized reactivity tests comparing acylation of small bulge loops with that of a larger hairpin loop for site-selective labeling within a folded motif

(A) Schematic of site-selective acylation of a designed "bulge loop tag" (BLT-A_3 or BLT-U_1) including U_5 hairpin and A_3 or U_1 bulge loop.

(B and C) CE analyses of acylation reactivity are shown for BLT-A_3 (B) and BLT-U1 (C); acylation reactivity is calculated by the peak area of corresponding RT stops from CE data. In both cases, the bulge loops show higher reactivity than a larger hairpin loop in the same RNA.

including hypothesized sugar conformation and catalytic effects (McGinnis et al., 2012). In our data, the widest reactivity variations do not appear to readily correlate purely with flexibility; for example, large loops (presumably more flexible) react less strongly than small ones, and unconstrained ssRNAs react relatively poorly as well. We suggest that a combination of

conformational constraints in small loops that enforce accessible or reactive conformations (either in backbone rotamers, sugar conformations, or both) along with dynamic flexibility that confers transient access together govern the reactivity of nucleotides in these folded RNAs. To shed more light on these contributions, it will be of special interest in the future to study in greater depth the dynamic effects of conformation on reactivity with fast- and slower reacting reagents.

Our identified sequence-generic reactivity patterns among RNA motifs could potentially help to improve structure-mapping interpretation, as well as the analysis of RNA-protein interactions. Folding algorithms commonly predict a suite of candidate structures containing different loop domains, and mining of prior data has been useful in more accurate calling of folds from acylation mapping (Cao and Xue, 2021; Ledda and Aviran, 2018); for example, a recent study analyzed data from 11 biological RNAs. The addition of 43 generic RNA sequences should help increase the diversity of datasets. In addition, our data have shown a wide range of reactivity in ssRNA, depending on nearest neighbors, suggesting an inherent sequence-dependent SHAPE reactivity difference, which may be useful to incorporate into future algorithms for better modeling of data in longer unpaired regions. Finally, our data may also be helpful in interpreting RNA-protein interactions by establishing additional RNA backgrounds and baselines for comparison. For example, our gel analysis of low-reactive ssRNA sequences (Figure 5B) resembles a classic "footprint" pattern but is actually not a footprint, given the lack of protein or other ligands. In the absence of these data, one might suspect a protein binding site if the experiment were performed in mixed biological media. Thus, it will be important to recognize this sequence variability as a comparative baseline for future analysis of interactions with ssRNA sequences.

We expect that our findings could be especially useful in the design of new methods for high-yield RNA labeling and conjugation. For example, the observation of highly reactive small bulge loops, and their surprisingly elevated reactivity relative to most other unpaired RNA nucleotides, suggests the interesting possibility of structurally localized RNA reaction. In future work, it may be possible to incorporate highly reactive and selective acylation tag motifs into RNAs of interest, as a strategy to enhance yield at selected sites. With functional groups included in some newer acylating reagents, such as the azide group in NAI-N_3, acyl adducts can be further substituted by fluorescent labels, by biotin, or other useful species (Velema and Kool, 2020).

Considering limitations and future directions of the current work, we note that the homogeneous loop sequences of the RNA library do not address questions of how specific sequences affect loop structure and reactivity. More information could be obtained by screening comprehensively varied loop sequences, especially in medium-sized loops, which may uncover specialized cross-loop interactions that expose certain 2'-OH groups and protect others. In addition, homogeneous *in vitro* probing in this simplified library does not address the effects of RNA-RNA interactions and protein-RNA interactions that occur in biological settings. Finally, testing a broader set of structurally varied acylation reagents, or reagents with a wider range of reactivity, may also prove instructive by providing a fuller understanding of steric and dynamic effects influencing the reaction and potentially yielding new reagents with altered selectivity as well.

SIGNIFICANCE

RNA acylation at 2'-OH groups is broadly useful in mapping folded structure and more recently has been adopted as a chemical strategy for high-yield conjugation of the biopolymer. To date, studies carried out to understand what structural features in RNA lead to localized reaction have been confined to limited sets of biological RNAs with their native sequences. This leaves open questions of whether the conclusions are general or are restricted to specific sequences of loops. This study addresses this issue by the use of libraries of "generic" folded RNAs containing loops with homogeneous sequences of all uridines or all adenosines. The study shows similar effects for two well-known classes of probing reagents, and thus, the results appear to be general.

This work reveals new insights into how folded structure helps determine the reactivity of certain RNA positions over others and yields multiple surprising findings. Among these findings are that nearest neighbors have a strong influence on reactivity and determine the reactivity of purely single-stranded sequences. Unexpectedly, we find that certain small bulge and internal loops are much more reactive than ssRNAs, despite high flexibility in the latter. Moreover, we show that small loops in general react more rapidly than do large loops. Finally, we show that, using this information, we can design small highly reactive acylation tags that permit localized reaction. Taken together, this systematic information of RNA acylation reactivity will inform efforts in refining RNA structure mapping and interaction and also can provide insights into future RNA designs for high-yield local labeling and conjugation.

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
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - *In vitro* RNA acylation
  - *In vitro* manual SHAPE analysis of representative folded structure
  - Deep sequencing library preparation
  - Bioinformatics for data analysis
  - Demonstration of designed single-stranded RNA
  - Demonstration of designed bulge loop tag (BLT) RNA
  - DNA template was purified by gel extraction kit (Qiagen)
  - DMS probing assay
  - SHAPE mapping comparison assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2022.05.005.

ACKNOWLEDGMENTS

The authors thank the US National Institutes of Health (GM127295 and GM130704) for support.

AUTHOR CONTRIBUTIONS

E.T.K. supervised all aspects of the project; L.X. and E.T.K. designed the research; L.X. and L.F. performed the experiments; L.X., L.F., and E.T.K. analyzed and interpretated data; L.X. and E.T.K. wrote the manuscript; and L.F. contributed to the writing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2022
Revised: March 16, 2022
Accepted: May 12, 2022
Published: June 2, 2022

REFERENCES

Boerneke, M.A., Ehrhardt, J.E., and Weeks, K.M. (2019). Physical and functional analysis of viral RNA genomes by SHAPE. Annu. Rev. Virol. *6*, 93–117. https://doi.org/10.1146/annurev-virology-092917-043315.

Bonneau, E., and Legault, P. (2014). Nuclear magnetic resonance structure of the III–IV–V three-way junction from the varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. Biochemistry *53*, 6264–6275. https://doi.org/10.1021/bi500826n.

Brown, R.F., Andrews, C.T., and Elcock, A.H. (2015). Stacking free energies of all DNA and RNA nucleoside pairs and dinucleoside-monophosphates computed using recently revised AMBER parameters and compared with experiment. J. Chem. Theory Comput. *11*, 2315–2328. https://doi.org/10.1021/ct501170h.

Busan, S., Weidmann, C.A., Sengupta, A., and Weeks, K.M. (2019). Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. Biochemistry *58*, 2655–2664. https://doi.org/10.1021/acs.biochem.8b01218.

Butcher, S.E., Allain, F.H.T., and Feigon, J. (1999). Solution structure of the loop B domain from the hairpin ribozyme. Nat. Struct. Biol. *6*, 212–216. https://doi.org/10.1038/6651.

Butcher, S.E., Dieckmann, T., and Feigon, J. (1997). Solution structure of a GAAA tetraloop receptor RNA. EMBO J. *16*, 7490–7499. https://doi.org/10.1093/emboj/16.24.7490.

Cao, J., and Xue, Y. (2021). Characteristic chemical probing patterns of loop motifs improve prediction accuracy of RNA secondary structures. Nucleic Acids Res. *49*, 4294–4307. https://doi.org/10.1093/nar/gkab250.

Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-ela, F., Varani, G., and Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. Nat. Struct. Biol. *9*, 375–380. https://doi.org/10.1038/nsb785.

Cordero, P., Kladwang, W., VanLang, C.C., and Das, R. (2012). Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. Biochemistry *51*, 7037–7039. https://doi.org/10.1021/bi3008802.

Danaee, P., Rouches, M., Wiley, M., Deng, D., Huang, L., and Hendrix, D. (2018). bpRNA: large-scale automated annotation and analysis of RNA secondary structure. Nucleic Acids Res. *46*, 5381–5394. https://doi.org/10.1093/nar/gky285.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. U S A *106*, 97–102. https://doi.org/10.1073/pnas.0806929106.

Desjardins, G., Bonneau, E., Girard, N., Boisbouvier, J., and Legault, P. (2011). NMR structure of the A730 loop of the Neurospora VS ribozyme: insights into the formation of the active site. Nucleic Acids Res. *39*, 4427–4437. https://doi.org/10.1093/nar/gkq1244.

Diener, J.L., and Moore, P.B. (1998). Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. Mol. Cell *1*, 883–894. https://doi.org/10.1016/S1097-2765(00)80087-8.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700. https://doi.org/10.1038/nature12756.

Fessler, A.B., Fowler, A.J., and Ogle, C.A. (2021). Directly quantifiable biotinylation using a water-soluble isatoic anhydride platform. Bioconjugate Chem. *32*, 904–908. https://doi.org/10.1021/acs.bioconjchem.1c00150.

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc. *11*, 273–290. https://doi.org/10.1038/nprot.2016.011.

Fürtig, B., Richter, C., Wöhnert, J., and Schwalbe, H. (2003). NMR spectroscopy of RNA. Chembiochem *4*, 936–962. https://doi.org/10.1002/cbic.200300700.

Ganser, L.R., Kelly, M.L., Herschlag, D., and Al-Hashimi, H.M. (2019). The roles of structural dynamics in the cellular functions of RNAs. Nat. Rev. Mol. Cell Biol. *20*, 474–489. https://doi.org/10.1038/s41580-019-0136-0.

Gopal, A., Zhou, Z.H., Knobler, C.M., and Gelbart, W.M. (2012). Visualizing large RNA molecules in solution. RNA *18*, 284–299. https://doi.org/10.1261/rna.027557.111.

Greenbaum, N.L., Radhakrishnan, I., Patel, D.J., and Hirsh, D. (1996). Solution structure of the donor site of a trans-splicing RNA. Structure *4*, 725–733. https://doi.org/10.1016/S0969-2126(96)00078-0.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The vienna RNA websuite. Nucleic Acids Res. *36*, W70–W74. https://doi.org/10.1093/nar/gkn188.

Habibian, M., Velema, W.A., Kietrys, A.M., Onishi, Y., and Kool, E.T. (2019). Polyacetate and polycarbonate RNA: acylating reagents and properties. Org. Lett. *21*, 5413–5416. https://doi.org/10.1021/acs.orglett.9b01526.

Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H., and Weeks, K.M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc. Natl. Acad. Sci. U S A *110*, 5498–5503. https://doi.org/10.1073/pnas.1219988110.

Jash, B., and Kool, E.T. (2022). Conjugation of RNA via 2′-OH acylation: mechanisms determining nucleotide reactivity. Chem. Commun. *58*, 3693–3696. https://doi.org/10.1039/D2CC00660J.

Jucker, F.M., and Pardi, A. (1995). Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. Biochemistry *34*, 14416–14427. https://doi.org/10.1021/bi00044a019.

Kadina, A., Kietrys, A.M., and Kool, E.T. (2018). RNA cloaking by reversible acylation. Angew. Chem. Int. Ed. *57*, 3059–3063. https://doi.org/10.1002/anie.201708696.

Kim, C.-H., and Tinoco, I. (2001). Structural and thermodynamic studies on mutant RNA motifs that impair the specificity between a viral replicase and its promoter11Edited by D. Draper. J. Mol. Biol. *307*, 827–839. https://doi.org/10.1006/jmbi.2001.4497.

Knorre, D.G., Pustoshilova, N.M., Teplova, M., and Shamovskiĭ, G.G. (1965). The production of transfer RNA acetylated by 2'-oxy groups. Biokhimiia (Moscow, Russia) *30*, 1218–1224.

Ledda, M., and Aviran, S. (2018). PATTERNA: transcriptome-wide search for functional RNA elements via structural data signatures. Genome Biol. *19*, 28. https://doi.org/10.1186/s13059-018-1399-z.

Lee, B., Flynn, R.A., Kadina, A., Guo, J.K., Kool, E.T., and Chang, H.Y. (2017). Comparison of SHAPE reagents for mapping RNA structures inside living cells. RNA *23*, 169–174. https://doi.org/10.1261/rna.058784.116.

Lescrinier, E.M.H.P., Tessari, M., van Kuppeveld, F.J.M., Melchers, W.J.G., Hilbers, C.W., and Heus, H.A. (2003). Structure of the pyrimidine-rich internal loop in the poliovirus 3′-UTR: the importance of maintaining pseudo-2-fold symmetry in RNA helices containing two adjacent non-canonical base-pairs. J. Mol. Biol. *331*, 759–769. https://doi.org/10.1016/S0022-2836(03)00787-3.

Lipsett, M.N. (1960). Evidence for helical structure in polyuridylic acid. Proc. Natl. Acad. Sci. U. S. A *46*, 445–446. https://doi.org/10.1073/pnas.46.4.445.

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. https://doi.org/10.1016/j.cell.2016.04.028.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. U S A *101*, 7287–7292. https://doi.org/10.1073/pnas.0401799101.

McGinnis, J.L., Dunkle, J.A., Cate, J.H.D., and Weeks, K.M. (2012). The mechanisms of RNA SHAPE chemistry. J. Am. Chem. Soc. *134*, 6617–6624. https://doi.org/10.1021/ja2104075.

McGinnis, J.L., Liu, Q., Lavender, C.A., Devaraj, A., McClory, S.P., Fredrick, K., and Weeks, K.M. (2015). In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. Proc. Natl. Acad. Sci. U S A *112*, 2425–2430. https://doi.org/10.1073/pnas.1411514112.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005). RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J. Am. Chem. Soc. *127*, 4223–4231. https://doi.org/10.1021/ja043822v.

Moore, P.B., and Steitz, T.A. (2003). The structural basis of large ribosomal subunit function. Annu. Rev. Biochem. *72*, 813–850. https://doi.org/10.1146/annurev.biochem.72.110601.135450.

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. https://doi.org/10.1021/ja0704028.

Nomura, Y., Kajikawa, M., Baba, S., Nakazato, S., Imai, T., Sakamoto, T., Okada, N., and Kawai, G. (2006). Solution structure and functional importance of a conserved RNA hairpin of eel LINE UnaL2. Nucleic Acids Res. *34*, 5184–5193. https://doi.org/10.1093/nar/gkl664.

Park, H.S., Kietrys, A.M., and Kool, E.T. (2019). Simple alkanoyl acylating agents for reversible RNA functionalization and control. Chem. Commun. *55*, 5135–5138. https://doi.org/10.1039/C9CC01598A.

Peritz, A.E., Kierzek, R., Sugimoto, N., and Turner, D.H. (1991). Thermodynamic study of internal loops in oligoribonucleotides: symmetric loops are more stable than asymmetric loops. Biochemistry *30*, 6428–6436. https://doi.org/10.1021/bi00240a013.

Popenda, L., Bielecki, L., Gdaniec, Z., and Adamiak, R.W. (2008). Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple. ARKIVOC *2009*, 130–144. https://doi.org/10.3998/ark.5550190.0010.311.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform. *11*, 129. https://doi.org/10.1186/1471-2105-11-129.

Richards, E.G. (1968). On the analysis of melting curves of stacked polynucleotides. Eur. J. Biochem. *6*, 88–92. https://doi.org/10.1111/j.1432-1033.1968.tb00423.x.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C., and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. Nature *250*, 546–551. https://doi.org/10.1038/250546a0.

Sharp, P.A. (2009). The centrality of RNA. Cell *136*, 577–580. https://doi.org/10.1016/j.cell.2009.02.007.

Smith, J.S., and Nikonowicz, E.P. (1998). NMR structure and dynamics of an RNA motif common to the spliceosome branch-point helix and the RNA-binding site for phage GA coat protein. Biochemistry *37*, 13486–13498. https://doi.org/10.1021/bi981558a.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20. https://doi.org/10.1038/nchembio.1131.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints *in vivo* decode RNA regulatory mechanisms. Nature *519*, 486–490. https://doi.org/10.1038/nature14263.

Thapar, R., Denmon, A.P., and Nikonowicz, E.P. (2014). Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. WIREs RNA *5*, 49–67. https://doi.org/10.1002/wrna.1196.

Tijerina, P., Mohr, S., and Russell, R. (2007). DMS footprinting of structured RNAs and RNA–protein complexes. Nat. Protoc. *2*, 2608–2623. https://doi.org/10.1038/nprot.2007.380.

Turner, D.H., and Mathews, D.H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res. *38*, D280–D282. https://doi.org/10.1093/nar/gkp892.

Ursuegui, S., Chivot, N., Moutin, S., Burr, A., Fossey, C., Cailly, T., Laayoun, A., Fabis, F., and Laurent, A. (2014). Biotin-conjugated N-methylisatoic anhydride: a chemical tool for nucleic acid separation by selective 2′-hydroxyl acylation of RNA. Chem. Commun. *50*, 5748–5751. https://doi.org/10.1039/C4CC01134A.

Velema, W.A., and Kool, E.T. (2020). The chemistry and applications of RNA 2′-OH acylation. Nat. Rev. Chem *4*, 22–37. https://doi.org/10.1038/s41570-019-0147-6.

Velema, W.A., Kietrys, A.M., and Kool, E.T. (2018). RNA control by photoreversible acylation. J. Am. Chem. Soc. *140*, 3491–3495. https://doi.org/10.1021/jacs.7b12408.

Wang, P.Y., Sexton, A.N., Culligan, W.J., and Simon, M.D. (2019). Carbodiimide reagents for the chemical probing of RNA structure in cells. RNA *25*, 135–146. https://doi.org/10.1261/rna.067561.118.

Weidmann, C.A., Mustoe, A.M., Jariwala, P.B., Calabrese, J.M., and Weeks, K.M. (2021). Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA. Nat. Biotechnol. *39*, 347–356. https://doi.org/10.1038/s41587-020-0709-7.

Weng, X., Gong, J., Chen, Y., Wu, T., Wang, F., Yang, S., Yuan, Y., Luo, G., Chen, K., Hu, L., et al. (2020). Keth-seq for transcriptome-wide RNA structure mapping. Nat. Chem. Biol. *16*, 489–492. https://doi.org/10.1038/s41589-019-0459-3.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2005). RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNAAsp transcripts. J. Am. Chem. Soc. *127*, 4659–4667. https://doi.org/10.1021/ja0436749.

Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., Giddings, M.C., and Weeks, K.M. (2009). Influence of nucleotide identity on ribose 2′-hydroxyl reactivity in RNA. RNA *15*, 1314–1321. https://doi.org/10.1261/rna.1536209.

Wu, Y., Shi, B., Ding, X., Liu, T., Hu, X., Yip, K.Y., Yang, Z.R., Mathews, D.H., and Lu, Z.J. (2015). Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. Nucleic Acids Res. *43*, 7247–7259. https://doi.org/10.1093/nar/gkv706.

Xiao, L., Habibian, M., and Kool, E.T. (2020). Site-selective RNA functionalization via DNA-induced structure. J. Am. Chem. Soc. *142*, 16357–16363. https://doi.org/10.1021/jacs.0c06824.

Xiao, L., Jun, Y.W., and Kool, E.T. (2021). DNA tiling enables precise acylation-based labeling and control of mRNA. Angew. Chem. Int. Ed. *n/a*. *60*, 26798–26805. https://doi.org/10.1002/anie.202112106.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415. https://doi.org/10.1093/nar/gkg595.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
2-(Azidomethyl)-3-pyridinyl]-1H-imidazol-1-yl-methanone (NAI-N3)	(Spitale et al., 2015)	N/A
1-Methyl-7-nitroisatoic anhydride (1M7)	Sigma-Aldrich	CAT# 908,401
Dimethylsulfate (DMS)	Sigma-Aldrich	CAT# D186309
5M NaCl	InvitrogenTM	CAT# AM9760G
1M MgCl2	InvitrogenTM	CAT# AM9530G
HEPES	Gibco™	CAT# 15630080
DMSO	ACROS Organics	CAT#B0532976
UltraPure DTT	InvitrogenTM	CAT# 15508013
NTP Set (100 mM Solution)	Thermo Scientific™	CAT# R0481
dNTP mix (10 mM each)	Thermo Scientific™	CAT# R0194
SYBR Green I nucleic acid gel stain (10,000×)	InvitrogenTM	CAT# S7567
SequaGel - UreaGel Concentrate	National diagnostics	CAT# EC-830
SequaGel - UreaGel Diluent	National diagnostics	CAT# EC-840
SequaGel - UreaGel Buffer	National diagnostics	CAT# EC-835
Ammonium Persulfate	Thermo Scientific™	CAT# 17874
N, N, N′, N′-Tetramethyl ethylenediamine	Sigma-Aldrich	CAT# 110732
Topvision agarose	Thermo Scientific™	CAT# R0491
10xTBE buffer	KD Medical	CAT# RGF-3330
RNaseOUT™ Recombinant Ribonuclease Inhibitor	InvitrogenTM	CAT# 10777019
RiboLock RNase inhibitor	Thermo Scientific™	CAT# EO0382
FastAP thermosensitive alkaline phosphatase	Thermo Scientific™	CAT# EF0652
DNase I	New England BioLabs	CAT# M0303S
T4 polynucleotide kinase	New England BioLabs	CAT# M0201S
T4 RNA ligase 1	New England BioLabs	CAT# M0204S
RNase cocktail enzyme mix	InvitrogenTM	CAT# AM2286
RNase H	Thermo Scientific™	CAT# EN0201
CircLigase II ssDNA ligase	Epicentre	CAT# CL9025K
Critical commercial assays		
Corning Costar Spin-X centrifuge filters, 0.45 μm	Millipore-Sigma	CAT# CLS8162
RNA clean-up and concentrator-5 column	Zymo Research	CAT# R1016
DNA clean-up and concentrator-5 column	Zymo Research	CAT# D4014
Amicon Ultra-0.5 Centrifugal Filter Unit	Millipore-Sigma	CAT# UFC500396
SuperScript™ III Reverse Transcriptase	InvitrogenTM	CAT# 18080044
Phusion high-fidelity (HF) PCR master mix	New England BioLabs	CAT# M0531S
Q5 hot start high fidelity PCR master mix	New England BioLabs	CAT# M0494S
HiScribe™ T7 Quick High Yield RNA Synthesis Kit	New England BioLabs	CAT# E2050S
MiniElute gel extraction Kit	QIAGEN	CAT# 28604

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
A-case RNA library, see Table S1	Integrated DNA Technologies	N/A
U-case RNA library, see Table S1	Integrated DNA Technologies	N/A
RNA linker for library preparation, see method details	Integrated DNA Technologies	N/A
RT-primer for library preparation, see method details	Integrated DNA Technologies	N/A
DNA primers for qPCR, see method details	Integrated DNA Technologies	N/A
Chimeric RNA (HP-4A-chimeric), see Table S1	Integrated DNA Technologies	N/A
Designed ssRNA (ss-ML), see Table S1	Integrated DNA Technologies	N/A
Primers for reverse transcription, see Table S1	Integrated DNA Technologies	N/A
DNA templates for transcription, see Table S1	Integrated DNA Technologies	N/A
Deposited data		
RNA-Seq data	This paper	GEO Accession# GSE199394
NMR structure of trinucleotide hairpin RNA	(Kim and Tinoco, 2001)	PDB# 1i4b
NMR structure of tetraloop hairpin RNA	(Butcher et al., 1999)	PDB# 1AFX
NMR structure of tetraloop hairpin RNA	(Nomura et al., 2006)	PDB# 2FDT
NMR structure of pentaloop hairpin RNA	(Jucker and Pardi, 1995)	PDB# 1RNG
NMR structure of hexaloop hairpin RNA	(Nomura et al., 2006)	PDB# 2FDT
NMR structure of 2x2 internal loop RNA	(Lescrinier et al., 2003)	PDB# 1N66
NMR structure of 3x1 internal loop RNA	(Desjardins et al., 2011)	PDB# 2L5Z
NMR structure of 3x2 internal loop RNA	(Butcher et al., 1997)	PDB# 1TLR
NMR structure of 4x2 internal loop RNA	(Collier et al., 2002)	PDB# 1KP7
NMR structure of single nucleotide bulge loop RNA	(Smith and Nikonowicz, 1998)	PDB# 17RA
NMR structure of single nucleotide bulge loop RNA	(Greenbaum et al., 1996)	PDB# 1SLP
NMR structure of single nucleotide bulge loop RNA	(Popenda et al., 2008)	PDB# 2K41
NMR structure of trinucleotide bulge loop RNA	(Diener and Moore, 1998)	PDB# 2A9L
NMR structure of three-way junction RNA	(Bonneau and Legault, 2014)	PDB# 2MTJ
Software and algorithms		
icSHAPE pipeline	(Flynn et al., 2016)	N/A
ImageJ	NIH	N/A
StepOnePlus Real-Time PCR System	Applied Biosystems	N/A
GraphPad Prism 7	GraphPad Prism Software, Inc.	N/A
BioRender	BioRender.com	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Eric T. Kool (kool@stanford.edu).

Materials availability

This study did not generate unique reagents.

Data and code availability

- Sequencing data reported in this paper has been deposited in Gene Expression Omnibus (GEO Accession# GSE199394). Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All data are generated from the datasets provided in the KRT.

METHOD DETAILS

In vitro RNA acylation

RNAs were chemically synthesized and purified by gel electrophoresis. For acylation conditions optimization, 1 μL 50 μM hp-4A-chimeric RNA (4A-tetraloop hairpin with a chimeric DNA tail for primer landing) wasin heated in metal-free water for 2 min at 95°C and was flash cooled on ice. 1.25 μL 8×SHAPE (800 mM HEPES, pH 7.5, 48 mM $MgCl_2$, 800 mM NaCl) buffer was added and the RNA solution was allowed to equilibrate at 37°C for 20 min. To this mixture, 1 μL of 10x acylating reagents (NAI-N_3 and 1M7) was added to a different 1x final concentration (NAI-N_3: 20mM–200mM; 1M7: 2–25mM). The reaction was permitted to continue at 37°C until the desired time (2–10min) and was quenched by DTT with 1.25 equivalent amount of acylating reagents at 37°C for 20 min. Acylated RNA was purified by ethanol precipitation and stored at −80°C for SHAPE analysis.

For library RNA acylation, each RNA in the library was first folded individually and then equally mixed with others in the set. Briefly, 1 μL 50 μM each RNA strand was heated in 0.75 μL metal-free water for 2 min at 95°C and flash-cooled on ice. For folding, 0.25 μL 8×SHAPE was added and incubated at 37°C for 20 min. The 22 strands of structured RNAs (25 μM each) were mixed to make a A-case or U-case structured RNA library. Next, in a final 10 μL acylation reaction, 2 μL 25 μM structured library RNA mixtures were added with 1 μL 10x acylating reagents (2 M NAI-N_3 or 0.25 M 1M7) in DMSO (+) or DMSO (−). The reaction was incubated at 37°C for 10 min and was quenched by DTT with 1.25 equivalent amount of acylating reagents at 37°C for 20 min. Acylated RNA libraries were purified by 3K-amicon centrifuge filter washing with RNase-free water. The purified and reacted RNAs solutions were aliquoted to ∼500ng per tube and were freeze-dried on a lyophilizer prior to deep-sequencing library preparation.

In vitro manual SHAPE analysis of representative folded structure

7 pmol acylated hp-4A-chimeric RNA was mixed with 7 pmol Reverse transcription (RT) primer and 0.5 μL dNTP mix (10 mM each, Invitrogen), and incubated for 5 min at 70°C, followed by by a step-down cooling (1° per 3s) to 4°C. Then 2 μL 5x First-Strand Buffer (Invitrogen), 0.5 μL 0.1 M DTT, 0.375 μL RNaseOUT and 0.625 μL Super Script III (200 U/μL, Invitrogen) were added to the final volume of 10 μL. The reaction was incubated with the following program: 25°C for 10 min, 42°C for 50 min, and 52°C for 50 min. After the reaction, 10 μL loading dye (8 M Urea, 0.05% Orange G, 0.05% Bromophenol blue) was added and the mixture was denatured at 96°C for 3 min and loaded on a denaturing 20% polyacrylamide gel (PAGE). Products were separated in a PAGE gel in 1x TBE (pH 8.3, Sigma Aldrich), 20 mA, ∼2 h. The cDNA gel was visualized by fluorescence imaging (Typhoon, GE Healthcare).

Deep sequencing library preparation

(1) RNA 3'-end repair: All RNA library samples (NAI-N_3, 1M7 and DMSO treated) were processed through a ends repair to resolve the 3'-end phosphates. Re-dissolve RNA in 10 μL of RNA end-repair mix and incubate at 37°C for 1 h. RNA end-repair mix: 1 μL 10x T4 PNK buffer (NEB), 1 μL 40U/μL RiboLock (Thermo Fisher), 1 μL FastAP 1 U/μL (Thermo), 2 μL T4 PNK 10,000 U/mL (NEB) and 5 μL RNase-free water.

(2) RNA 3'-end ligation: 15.875 μL of the following 3' RNA ligation mixture was added to above 10 μL completed end-repair reaction. Mix by flickering and incubate at 25°C for 4 h in a thermocycler. 3' RNA ligation mixture: 1.5 μL 10x RNA ligase buffer, 0.625 μL DTT, 200 mM (Fresh), 2.5 μL 10 μM preadenylylated and 3'-ddC blocked RNA linker, 3.75 μL T4 RNA ligase 1 (10,000 U/mL), 7.5 μL PEG8000, 50% (v/v). After ligation, 24.2 μL of water was added to reach a total volume of 50 μL and RNAs were purified with a Zymo RNA clean&concentrator 5-colume. Freeze-dry samples on a lyophilizer.

pre-adenylylated and 3'-ddC blocked RNA linker:
/5rApp/AGATCGGAAGAGCGGTTCAG/3ddC/ (IDT, PAGE purified)

(3) RNA size selection: RNA samples were resuspended in 7 μL of gel loading buffer II and flicked to mix. 12% (w/v) UreaGel denaturing PAGE was pre-ran at 25 W for 3 min then 50 W for 5 min. RNA samples were loaded and ran at 20 W for ∼6 min. The gel was stained in SYBR Gold mix for 3 min at rt and imaged on a blue light transilluminator. Ligated RNAs ranging between 10 and 35nt (30nt-65nt with the 3' adaptor ligated) were excised. Gel slices were crushed through a 0.75 mL tube nested in a 1.5 mL tube by centrifugation. 250 μL water was added and was heated at 67°C for 15 min with shaking to elute RNAs twice. The gel slurry was transferred into a Spin-X column with a cut-tip p1000 and centrifuged at 6500g for 1 min at 25°C (final

eluate ~500 μL). 500 μL eluate was then transferred to 0.5-mL 3K Amicon column and centrifuge at 13500g for 15 min at rt to concentrate samples to ~100 μL. Next, RNA samples were further purified by Zymo RNA clean&concentrator-5 column and freeze-dried on a lyophilizer.

(4) Reverse transcription for cDNA synthesis: RNA samples were dissolved in 11 μL water and 1 μL of 10 μM RT primer was added. Solutions were mixed by flicking and moved to 200-μL PCR tubes. Samples were heated to 70°C for 5 min, cooled to 25°C by stepping down 1°C every 3 s (45 steps), holding at 25°C. To each solution, 8 μL of RT enzyme mix was added and heated to 25°C for 6 min, 42°C for 10 min, 52°C for 60 min, holding at 4°C. RT enzyme mix: 4 μL of 5× First Strand Buffer, 0.75 μL of RiboLock, 40U/μL, 1 μL 100 mM DTT and 1.25 μL of SuperScript III. Each reaction library used an RT primer containing a unique barcode to be identified after sequencing.

RT primer /5phos/DDDNNAACCNNNNAGATCGGAAGAGCGTCGTGGA/iSp18/GGATCC/iSp18/TACTGAACCGC; D = A/G/T and N = A/T/G/C are used to discriminate PCR duplicates. 'AACC' is the specific experimental barcode)

(5) RNA digestion: The above reverse transcribed cDNA samples were heated at 70°C for 15 min to deactivate RNase inhibitor and were mixed with 1 μL of 5U/μL RNase H and 1 μL of RNase cocktail enzyme mix. The mixtures were incubated at 37°C for 70 min. The reactions were then purified with DNA Zymo DNA clean&concentrator-5 column to recover single-stranded DNA≥ 20 nt with the following procedures: add 350 μL (7x volume) of DNA binding buffer and mix by vortexing. Add 350 μL (1x volume) of 100% ethanol and vortex to mix. Wash columns according to the manufacturer's protocol. Elute cDNA twice with 10 μL of water (final 20 μL) and freeze-dry on a lyophilizer.

(6) cDNA purification: cDNA was resuspended in 7 μL Gel loading buffer II and flicker to mix. 12% (w/v) denaturing PAGE was pre-run at 25 W for 3 min, then 50 W for 5 min. Samples were heated for 2 min at 95°C and placed on ice for 1 min cDNA samples were then loaded and run at 20W for ~18min. Gels were stained in SYBR Gold mix for 3 min at rt and the bands >60 nt were cut. cDNA was extracted as in RNA size selection (see above). 500 μL cDNA eluate was concentrated by 10K Amicon column and was purified with Zymo DNA clean&concentrator-5 column. The purified cDNA samples (~16 μL) were transferred to PCR tubes for the next step.

(7) cDNA circularization: 4 μL of circularization reaction mixture was added to above cDNA samples and the mixtures were incubate at 60°C for 3 h. Circularization reaction mixture: 2 μL 10x CircLigase II buffer, 1 μL 50 mM $MnCl_2$, 1 μL Circligase II. The circularized cDNA was then purified with Zymo DNA clean&concentrator-5 column and the eluant (~20 μL) was transferred to qPCR tube.

(8) Library amplification: 21.4 μL of qPCR reaction mix was added to the above circularized cDNA sample. The qPCR reaction was performed with program as below and amplification was tracked in real time and stopped at the linear part of the amplification curve. Amplified DNA samples were purified with Zymo DNA clean&concentrator-5 column by adding 5x DNA binding buffer. The eluant was freeze-dried on a lyophilizer and carried on to the size selection step.

qPCR reaction mix:

Component	Amount (μL)	Final concentration
Phusion HF PCR master mix, 2x	20	1x
SYBR Green I, 25x	0.4	0.25x
Solexa PCR primer, 10 μM (10 μM forward and 10 μM reverse)	1.0	0.5 μM
Total	21.4/reaction	

P5-Solexa PCR primer (IDT, PAGE purified):
5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT-3′
P3-Solexa PCR primer (IDT, PAGE purified):
5′-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCT-3′
Program:

Cycle number	Denature	Anneal	Extend
Cycle number	Denature	Anneal	Extend
1	98°C, 45 s	–	–
2 – n	98°C, 15 s	65°C, 20 s	72°C, 60 s

(9) library size selection: 3% low melting point agarose gels (Topvision agarose) were prepared in 1% TBE containing 1x SybrGold. PCR DNA samples were dissolved in 9 μL water and mixed with 3 μL 6xOrangeG dye. After sample loading, the gel was run at 70 V for 1.5 h. PCR DNAs ranging between ~120 and 155nt were cut and isolated with MiniElute gel extraction Kit (QIAGEN). The size-selected samples were quantified by Bioanalyzer. Samples reacting under different conditions were pooled together for multiplexing to proceed to deep sequencing.

Bioinformatics for data analysis

Paired-end sequencing was performed on an Illumina HiSeq sequencer, yielding approximately 45 million raw reads. We utilized the established icSHAPE pipeline (Flynn et al., 2016) to analyze our sequencing data with some modifications in parameter setting (used commands are showed below) Briefly, the raw sequencing data was processed by demultiplexing according to the barcode, collapsing to remove PCR duplicates, primer and linker trimming. RNA library index files were built and trimmed reads were mapped to sequences in each RNA library. The absolute RT-stop frequencies for each condition were calculated followed the scripts in the pipeline. We defined reactivity score (R) as the subtraction of background reverse transcription stops (DMSO libraries) adjusted by the background base density from reverse transcription stops of the acylated libraries adjusted by the sample base density:

$$R = (\text{RT-Stop}_{\text{sample}} / \text{base\_density}_{\text{sample}}) - (\text{RT-Stop}_{\text{DMSO}} / \text{base\_density}_{\text{DMSO}})$$

Bioinformatics commands

1. Demultiplex the sequencing data according to the barcode of each sub-library.

$ splitFastq.pl -U seq1_20210825_CKDL210018846-1a_HGKLWBBXX_L7.fastq -l GGTT:A-200-NAI-1::TTGT:A-200-NAI-2::TGGC:A-DMSO-1::GGTC:U-200-NAI-1::AATC:U-200-NAI-2:: AGGA:U-DMSO-1::GTTA:A-25-1M7-1::CTTC:A-25-1M7-2::ATGC:A-DMSO-2::ACGA:U-25-1M7-1::TAAG:U-25-1M7-2::CGCA:U-DMSO-2::others:unmatched -b 5:4 -d library_split_seq1 -s hiseq_barcode.stat

2. Use FastQC to check the quality of the reads from each sub-libraries:

$ fastqc -o out_directory LIB_x.fastq

3. Collapse the subreads for each library to remove PCR duplicates:

$ readCollapse.pl -U LIB_x.fastq -o LIB_x.rmdup.fastq -f LIB_x.fa

4. Trim the adapter and barcode sequences from reads

$ trimming.pl -U LIB_x.rmdup.fastq -o LIB_x.trimmed.fastq -l 13 -t 0 -c phred33 -a adapter.fa -m 3

5. Map the reads to sequences in RNA library

$ bowtie2-build RNA-A-case.fa transcripts
$ bowtie2 -U LIB_x.trimmed.fastq -S LIB_x.sam -x /directory that stores my transcriptome-index files named as "transcripts.*"/transcripts –non-deterministic –time

6. Estimate the abundance of identified transcripts

$ estimateRPKM.pl -i LIB_x.sam -o LIB_x.rpkm

7. Calculate the RT stop frequencies in each transcript

$ calcRT.pl -i LIB_x.sam -o LIB_x.rt -r LIB_x.rpkm -c 1

8. Analysis of the RT stop frequencies and calculate the reactivity score by

$$R = (\text{RT-Stop}_{\text{sample}} / \text{base\_density}_{\text{sample}}) - (\text{RT-Stop}_{\text{DMSO}} / \text{base\_density}_{\text{DMSO}})$$

Demonstration of designed single-stranded RNA

(1) Acylation reaction: 150 pmol ssRNA was first heated to 95°C for 5 min and flash-cooled on ice. HEPES buffer was added to 1x final concentration (100 mM HEPES, pH 7.5, 6 mM $MgCl_2$, 100 mM NaCl) and the mixture was equilibrated at 37°C for 20 min. NAI-N_3 in DMSO was then added to a different final concentration ranging from 10 mM to 200 mM. The total volume of acylation reaction is 10 μL with 10% DMSO. Reaction was incubated at 37°C for 10 min and then quenched by 1.25 equivalent DTT

with further 20 min incubation at 37°C. Acylated ssRNA was purified by ethanol precipitation and further analysis by gel after reverse transcription.

(2) Reverse transcription and gel analysis: 10 pmol ssRNA was mixed with 10 pmol Reverse transcription (RT) primer and 0.5 μL dNTP mix (10 mM each, Invitrogen), and incubated for 5 min at 70°C, followed by by a step-down cooling (0.2° per s) to 25°C, putting on ice. Then 2 μL 5x First-Strand Buffer (Invitrogen), 1 μL 0.1 M DTT, 0.5 μL RNaseOUT and 0.5 μL Super Script II (200 U/μL, Invitrogen) were added to the final volume of 10 μL. The reaction was incubated with the following program: 25°C for 10 min, 42°C for 50 min, and 52°C for 50 min. After the reaction, 10 μL loading dye (8 M Urea, 0.05% Orange G, 0.05% Bromophenol blue) was added and the mixture was denatured at 96°C for 3 min and loaded on a denaturing 20% polyacrylamide gel (PAGE). Products were separated in a PAGE gel in 1x TBE (pH 8.3, Sigma Aldrich), 20 mA, ~2 h. The cDNA gel was visualized by fluorescence imaging (Typhoon, GE Healthcare).

Demonstration of designed bulge loop tag (BLT) RNA

(1) PCR of DNA template

Component	Amount (μL)	Final concentration
Q5 Hot start high-fidelity master mix, 2x	12.5	1x
10 μM forward primer	1.25	0.5 μM
10 μM reverse primer	1.25	0.5 μM
DMSO	0.75	3%
template		~5 ng
Total	25/reaction	

The reaction was performed in thermocycler following the program below:

Step	Initial Denature	Anneal	Extend	Final extension	Hold
1	98°C, 30 s	–	–		
2 (30 Cycles)	98°C, 10 s	65°C, 30 s	72°C, 30 s		
3				72°C, 2min	
4					4°C

DNA template was purified by gel extraction kit (Qiagen)

(2) *In vitro* transcription of BLT RNA: RNAs were synthesized by *in vitro* transcription using HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB), following the manufactory's protocol. The reaction was incubated at 37°C overnight. Transcribed RNAs were purified by ethanol precipitation.

(3) Acylation reaction: 50 pmol BLT RNA (BLT-A_3 / BLT-U_1) was first annealed with 80 pmole complementary DNA to protect capillary electrophoresis handle in 50 mM NaCl buffer. The RNA-DNA complex was heat at 75°C for 3 min; then cooled to 25°C by stepping down at 0.2°C/s and put on ice after finishing. HEPES buffer was added to 1x final concentration (100 mM HEPES, pH 7.5, 6 mM $MgCl_2$, 100 mM NaCl) and the mixture was equilibrated at 37°C for 20 min. NAI-N_3 in DMSO was then added to a final concentration of 50 mM. The total volume of acylation reaction is 10 μL with 10% DMSO. Reaction was incubated at 37°C for 10 min and then quenched by 1.25 equivalent DTT with further 20 min incubation at 37°C. Acylated BLT RNA with DNA protector was purified by Zymo RNA clean-up and concentrator-5 column. 5 μL DNase I and 5 μL 10x Reaction buffer were added to a final volume of 50 μL. The mixture was incubated at 37°C for 1 h to completely digest DNA protectors. RNA was purified again with Zymo clean-up and concentrator-5 column and further analysis by capillary electrophoresis (CE) after reverse transcription.

(4) Reverse transcription and capillary electrophoresis: 1.5 pmol Acylated BLT RNA was mixed with 2 pmol Reverse transcription (RT) primer and 0.25 μL dNTP mix (10 mM each), and incubated for 5 min at 65°C, chilling on ice. Then 2 μL 5x First-Strand Buffer, 1 μL 0.1 M DTT, 0.5 μL RNaseOUT and 0.5 μL Super Script III (200 U/μL) were added to the final volume of 10 μL. The reaction was incubated with the following program: 25°C for 10 min, 52°C for 50 min, 55°C for 50 min, and hold at 4°C. Different fluorophore-labeled RT primers were used for capillary electrophoresis shown as below:

	CE-mix-1	CE-mix-2
Channel 1: FAM	BLT RNA - NAI-N_3 treated	BLT RNA - acylated
Channel 2: HEX	BLT RNA - DMSO treated	BLT RNA - ddG
Channel 3: TAMRA	BLT RNA - ddC	BLT RNA - ddC

1 μL of each RT sample with each fluorophore was added together as a 3 μL mixture for CE analysis. Capillary electrophoresis was carried out by ABI 3130xl Genetic Analyzer.

DMS probing assay

DMS was carefully diluted threefold (dx3) in ethanol and the reactions were set up as follows:

Component	Amount (μL)	Final concentration
50 μM RNA	1 μL	2 μM
Water	19.875	
95°C 5min, chill on ice		
8xHEPES SHAPE buffer	3.125 μL	1x
Total = 24 μL, 37°C incubate for 30min		
dx3 DMS/ethanol stock	1 μL	1.34% (vol/vol)
Total = 25 μL		

The reaction was performed at 25°C for 1 min and then quenched by adding 475 μL 50% b-mercaptoethanol/H_2O, incubating at 25°C for 5 min. The products were purified by 3k-Amicon filter, washing with water 6 times. Methylated hairpin RNAs were then ligated to DNA linker with the same procedure as above. Reverse transcription was performed as follows: 4 pmol methylated RNA was mixed with 5 pmol Reverse transcription (RT) primer and 0.5 μL dNTP mix (10 mM each, Invitrogen), and incubated for 5 min at 70°C.

Component	Amount (μL)	Final concentration
2 μM RNA	2 μL	4 pmole
10 μM Cy5-primer	0.5 μL	5 pmole
10 mM dNTP or ddTTP/dNTP mix	0.25 μL	
Water	3.65 μL	
70°C 5min, followed by by a step-down cooling (1° per 3s) to 4°C.		
First strand buffer, 5x	2 μL	1x
0.1 M DTT	1 μL	
RNaseOUT, 40 U/μL	0.3 μL	
Super-Script III	0.3 μL	
Total = 10 μL		

Reactions were heated to 25°C for 10 min, 42°C for 10 min, 52°C for 60 min; hold at 4°C. After the reaction, 10 μL loading dye (8 M Urea, 0.05% Orange G, 0.05% Bromophenol blue) was added and the mixture was denatured at 96°C for 3 min and loaded on a denaturing 20% polyacrylamide gel (PAGE). Products were separated on a PAGE gel in 1x TBE (pH 8.3, Sigma Aldrich), 20 mA, ~2 h. The cDNA gel was visualized by fluorescence imaging (Typhoon, GE Healthcare).

SHAPE mapping comparison assay

Component	Amount (μL)	Final concentration
50 μM RNA	1 μL	5 μM
Water	5.75 μL	
95°C 5min, chill on ice		
8xHEPES SHAPE buffer or water	1.25 μL	1x
37°C incubate for 30min		
1 M NAI-N_3	2 μL	200 mM
Total = 10 μL		

The reaction was performed at 37°C for 10 min and then quenched by adding 12.5 μL 0.2M DTT, incubating at 37°C for 20min. The reactions were purified by 3k-Amicon filter, washing with water 6 times. Acylated hairpin RNAs were then ligated with DNA linker and performed reverse transcription and gel analysis with the same procedure as above.

QUANTIFICATION AND STATISTICAL ANALYSIS

The methods of statistical analysis are provided in Method Details and Figure Legends.