A single natural RNA modification can destabilize a U•A-T-rich RNA•DNA-DNA triple helix

CHARLOTTE N. KUNKLER,1 GRACE E. SCHIEFELBEIN,1 NATHAN J. O’LEARY,1 PHILLIP J. MCCOWN,2 and JESSICA A. BROWN1

1Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA
2Michigan Medicine, Department of Internal Medicine, Division of Nephrology, University of Michigan, Ann Arbor, Michigan 48109, USA

ABSTRACT

Recent studies suggest noncoding RNAs interact with genomic DNA, forming RNA•DNA-DNA triple helices, as a mechanism to regulate transcription. One way cells could regulate the formation of these triple helices is through RNA modifications. With over 140 naturally occurring RNA modifications, we hypothesize that some modifications stabilize RNA•DNA-DNA triple helices while others destabilize them. Here, we focus on a pyrimidine-motif triple helix composed of canonical U•A-T and C•G-C base triples. We employed electrophoretic mobility shift assays and microscale thermophoresis to examine how 11 different RNA modifications at a single position in an RNA•DNA-DNA triple helix affect stability: 5-methylcytidine (m5C), 5-methyluridine (m5U or rT), 3-methyluridine (m3U), pseudouridine (Ψ), 4-thiouridine (s4U), N6-methyladenosine (m6A), inosine (I), and each nucleobase with 2′-O-methylation (Nm). Compared to the unmodified U•A-T base triple, some modifications have no significant change in stability (Um•A-T), some have ∼2.5-fold decreases in stability (m5U•A-T, Ψ•A-T, and s4U•A-T), and some completely disrupt triple helix formation (m3U•A-T). To identify potential biological examples of RNA•DNA-DNA triple helices controlled by an RNA modification, we searched RMVar, a database for RNA modifications mapped at single-nucleotide resolution, for lncRNAs containing an RNA modification within a pyrimidine-rich sequence. Using electrophoretic mobility shift assays, the binding of DNA-DNA to a 22-mer segment of human lncRNA Al157886.1 was destabilized by ∼1.7-fold with the substitution of m5C at known m5C sites. Therefore, the formation and stability of cellular RNA•DNA-DNA triple helices could be influenced by RNA modifications.

Keywords: long noncoding RNA; RNA modification; triple helix

INTRODUCTION

Pyrimidine-motif RNA and DNA triple helices were shown to form in vitro over 60 yr ago (Felsenfeld and Rich 1957; Felsenfeld et al. 1957; Lipsett 1964; Riley et al. 1966; Morgan and Wells 1968). In such structures, a single strand of pyrimidine-rich RNA or DNA binds in a parallel orientation along the major groove of a purine-rich double-strand (ds) of Watson–Crick base-paired RNA or DNA, forming stacked base triples (Amott and Bond 1973; Amott and Selsing 1974; Amott et al. 1976). Since then, pyrimidine-motif triple helices, defined herein as three or more consecutive base triples, have been structurally validated in RNAs across all domains of life (Brown 2020). However, the largest pool of triple helices in vivo may be RNA•DNA-DNA triple helices (R•D-D, where “•” and “−” represent Hoogsteen and Watson-Crick interactions, respectively), which form between noncoding RNAs (ncRNAs) and genomic DNA (gDNA). Many proposed ncRNA•gDNA triple helices form between long noncoding RNA (lncRNA) and promoter regions of gDNA to regulate gene expression by recruiting chromatin-modifying enzymes (Sentürk Cetin et al. 2019). For example, the lncRNA Fendrr (Fetal lethal noncoding developmental regulatory RNA) forms a pyrimidine-motif triple helix within two different gDNA regions to silence Foxf1 and Pitx2 through the recruitment of chromatin-modifying complexes PRC2 and TrxG/MLL, respectively (Grote and Herrmann 2013; Grote et al. 2013). Additionally, the lncRNA Khsps1 (antisense to SPHK1) recruits the chromatin-modifying enzymes E2F1 and p300 to express SPHK1 (sphingosine kinase 1) (Postepska-Igielska et al. 2015; Blank-Giwojna et al. 2019). Considering that there are at least 27,000 genes encoding lncRNAs in human gDNA, the R•D-D triple helix could represent a
Large percentage of the triple helices in human cells and could be a general mechanism for RNA to regulate gene expression (Hon et al. 2017; Soibam 2017; Sentürk Cetin et al. 2019).

We recently determined the relative stability of 16 different base triples at a single position in an R•D-D triple helix (Kunkler et al. 2019). Though the most stable base triples for this triple helix are the canonical U•A-T and C•G-C base triples, other noncanonical base triples also allow for the formation of the triple helix (Kunkler et al. 2019). However, the effect of RNA modifications at a single position within an R•D-D triple helix has not yet been systematically studied. Over 140 different modifications of cellular RNA are known, with at least 12 identified in human IncRNAs (McCown et al. 2020; Yang et al. 2020b). One primary function of RNA modifications is to alter the stability of RNA secondary and tertiary structures (McCown et al. 2020). Therefore, RNA modifications may function as “switches” to regulate the formation of R•D-D triple helices in vivo and, in turn, regulate gene expression. Some modifications have been shown to change the stability of DNA•DNA-DNA (D•D-D) and RNA•RNA•RNA (R•R-R) triple helices in vitro, suggesting a stability change for R•D-D triple helices (Lee et al. 1984; Povsic and Dervan 1989; Xodo et al. 1991; Shimizu et al. 1992; Wang and Kool 1995; Wang et al. 1997; Zhou et al. 2013; Jacob et al. 2017). For example, 5-methylcytidine (m5C), one of the most abundant modifications in ncRNAs, was previously shown to stabilize D•D-D triple helices but destabilize R•R-R and R•D-R triple helices (Lee et al. 1984; Povsic and Dervan 1989; Xodo et al. 1991; Wang et al. 1997; Jacob et al. 2017). Another study showed that R•R-R triple helices are destabilized when the entire RNA third strand is 2′-O-methylated, while R•D-D triple helices are stabilized by an RNA third strand composed entirely of 2′-O-methylated uridines or cytidines (i.e., Um•A-U/T and Cm•G-C, respectively) (Shimizu et al. 1992; Wang and Kool 1995; Zhou et al. 2013). Because R•D-D triple helices have been implicated to play a role in transcription regulation for a variety of genes via ncRNA•gDNA interactions, the effect of RNA modifications on R•D-D triple helix formation is an exciting possibility to explore (Li et al. 2016).

For this study, 11 naturally occurring RNA modifications were chosen due to their commercial availability and, for most, their presence in human IncRNAs (underlined modifications have not been detected): 2′-O-methylcytidine (Cm), 5-methylcytidine (m5C), 2′-O-methyluridine (Um), 5-methyluridine (m5U; or ribothymidine, rT), 3′,5′-methyluridine (m3U), pseudouridine (Ψ), 4-thiouridine (s4U), 2′-O-methyladenosine (Am), N′-methyladenosine (m6A), inosine (I), and 2′-O-methylguanosine (Gm). To date, m6A has not been identified in human IncRNAs, though it is naturally occurring in human 28S ribosomal RNA (rRNA) (Taoka et al. 2018). Further, we expect a large destabilization of the R•D-D triple helix containing an m6A•X-Y base triple due to the methyl group disrupting Hoogsteen interactions, resulting in a destabilized R•D-D triple helix that serves as a control. s4U has not been identified in human RNAs, but it is used in RNA–protein, RNA–RNA, and RNA–DNA cross-linking experiments; therefore, the effect on R•D-D triple helix stability may be relevant within an experimental setting (Favre et al. 1986, 1998; Debreuil et al. 1991; Saintomé et al. 1997). Altogether, each of these modifications has been shown to either increase or decrease the stability of other nucleic acid structures, so it is likely that they also affect R•D-D triple helix stability (Lee et al. 1984; Huff and Topal 1987; Povsic and Dervan 1989; Xodo et al. 1991; Shimizu et al. 1992; Wang and Kool 1995; Mills et al. 1996; Kumar and Davis 1997; Wang et al. 1997; Zhou et al. 2013; Jacob et al. 2017; McCown et al. 2020). For each of the 11 RNA modifications, a single base triple in the R•D-D triple helix was varied and the relative stability of the triple helix was measured at a neutral pH using both electrophoretic mobility shift assays (EMSA) and microscale thermophoresis (MST). In general, our R•D-D triple helix with single-site RNA modifications showed no change, minor destabilization, or complete disruption of the triple helix, suggesting some RNA modifications, such as m5C in human IncRNA AL157886.1, could perturb the formation of R•D-D triple helices in vivo as a mechanism to regulate gene expression.

RESULTS AND DISCUSSION

Most RNA modifications destabilize an R•D-D triple helix

A physiologically relevant three-strand construct was previously used to determine the relative stability of 16 different unmodified R•D-D base triples at a single position using EMSA, concluding that the most stable base triples are the canonical U•A-T and C•G-C base triples (Kunkler et al. 2019). Herein, using the same parent construct and experimental setup, we tested the relative stability of 11 different RNA modifications (Supplemental Fig. S1) within Z•A-T and Z•G-C base triples (where Z = Cm, m5C, Um, m3U, m6A, Ψ, s4U, Am, m6A, I, Gm) to directly compare the binding to the previously examined unmodified base triples (Kunkler et al. 2019).

First, we examined the modified Z•A-T base triples (Fig. 1A). EMSAs were performed, whereby a shift of [32P]-radiolabeled dsDNA to a larger R•D-D complex was observed with increasing concentrations of the RNA strand (Fig. 1B). K_D,EMSA values were determined for each of the 11 modified base triples (Table 1; Fig. 1C,D; Supplemental Table S1), with Um•A-T exhibiting the tightest binding at 132 ± 8 nM and seven modified base triples completely disrupting the triple helix under our conditions: Cm•A-T, m5C•A-T, m3U•A-T, Am•A-T, m6A•A-T, I•A-T, and Gm•A-T (Table 1). It is perhaps not surprising that these modified noncanonical Z•A-T base triples

www.rnajournal.org 1173
completely destabilize the triple helix (Table 1; Fig. 1D), as our previous study showed that unmodified noncanonical Z•A-T base triples also had no observed triple helix formation (Kunkler et al. 2019). Additionally, the m₅U•A-T base triple completely destabilizes our triple helix, likely due to the N⁳-methyl group disrupting Hoogsteen base pairing. Three U-modified Z•A-T base triples (m₅U•A-T, Ψ•A-T, and s₄U•A-T) destabilize the triple helix to ~0.4 the stability of the unmodified U•A-T base triple (Table 1; Fig. 1D). Though we show a single m₅U•A-T base triple stabilizes an R•D-D triple helix, a previous study used a UV thermal denaturation assay to show that multiple m₅U•A-m₅U base triples stabilized an R•R-R and an R•D-R triple helix, likely due to the increased base stacking ability of m₅U base over uracil (Wang et al. 1997). To the best of our knowledge, Ψ and s₄U have not been studied in the context of an R•D-D triple helix, though both poly(Ψ) and poly(s₄U) can form R•R-R triple helices with poly(A) in the same 2:1 ratio as poly(U):poly(A) triple helices (Felsenfeld and Rich 1957; Felsenfeld et al. 1957; Simuth et al. 1970). It should be noted that even though s₄U is not naturally occurring in humans, it is used in biochemical techniques to increase the UV-crosslinking efficiency without greatly altering the chemical properties of the RNA. However, the 2.8-fold decrease in our R•D-D triple helices indicates that the addition of s₄U could lead to changes in R•D-D formation within these experiments. Herein, the Um•A-T base triple showed the highest stability with a Kₑ,EMSA of 132 ± 8 nM, minimally stabilizing the triple helix to ~1.4—the stability of the unmodified U•A-T base triple (Table 1; Fig. 1D). Previous studies have also shown stabilizing effects for the 2′-O-methylation of the RNA third
strand within an R•D-D triple helix via UV thermal denaturation studies and suggest the stabilization is due to the locking of the RNA into the 3′-endo conformation (Shimizu et al. 1992; Wang and Kool 1995). It should be noted that the previous studies had the third strand completely 2′-O-methylated (Shimizu et al. 1992; Wang and Kool 1995). Therefore, the minimal stabilizing effect measured herein might be greater with additional 2′-O-methyl nucleotides in the RNA third strand.

Using the same experimental setup, we determined \( K_{D,EMSA} \) and \( K_{D,MST} \) values for 11 different modifications within an R•D-D triple helix.

**TABLE 1.** \( K_{D,EMSA} \) and \( K_{D,MST} \) values for 11 different modifications within an R•D-D triple helix

| RNA Base (Z) | Z•A-T Base Triples | Z•G-C Base Triples |
|-------------|-------------------|-------------------|
|             | EMSA | MST | EMSA | MST |
| C           | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 165 ± 18 \( ^{\text{c}} \) | 1.0 | 128 ± 9 \( ^{\text{c}} \) | 1.0 |
| Cm          | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 440 ± 125 | 2.7 | 213 ± 16 | 1.7 |
| m\(^{\text{C}}\) | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 522 ± 132 | 3.2 | 268 ± 45 | 2.1 |
| U           | 187 ± 25 \( ^{\text{c}} \) | 1.0 | 109 ± 10 | 1.0 | 239 ± 14 \( ^{\text{c}} \) | 1.4 | 148 ± 73 | 1.2 |
| Um          | 132 ± 8 | 0.7 | 126 ± 17 | 1.2 | 299 ± 83 | 1.8 | 141 ± 22 | 1.1 |
| m\(^{\text{U}}\) | 482 ± 132 | 2.6 | 163 ± 31 | 1.5 | 310 ± 38 | 1.9 | 1,129 ± 518 | 8.8 |
| m\(^{\text{U}}\) | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 1,056 ± 235 | 6.4 | 1,326 ± 264 | 10 |
| w           | 443 ± 68 | 2.4 | 144 ± 11 | 1.3 | 329 ± 34 | 2.0 | 78 ± 6 | 0.6 |
| s\(^{\text{U}}\) | 528 ± 86 | 2.8 | 269 ± 32 | 2.5 | 855 ± 173 | 5.2 | 937 ± 161 | 7.3 |
| A           | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 207 ± 26 \( ^{\text{c}} \) | 1.3 | 175 ± 61 | 1.4 |
| Am          | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 428 ± 94 | 2.6 | 306 ± 68 | 2.4 |
| m\(^{\text{A}}\) | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 251 ± 22 | 1.5 | >50,000 \( ^{\text{a}} \) | >390 |
| i           | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 445 ± 50 | 2.7 | 359 ± 111 | 2.7 |
| G           | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 359 ± 33 \( ^{\text{c}} \) | 2.2 | >50,000 \( ^{\text{a}} \) | >390 |
| Gm          | >50,000 \( ^{\text{a}} \) | >270 | >50,000 \( ^{\text{a}} \) | >460 | 556 ± 75 | 3.4 | 572 ± 143 | 4.5 |

\(^{\text{a}}\)Reported \( K_{D,EMSA} \) and \( K_{D,MST} \) values are the average of at least three independent replicates and the associated standard deviation. Here, green shading corresponds to a low \( K_D \) value, yellow to an intermediate \( K_D \) value, and red represents no observed binding.

\(^{\text{b}}\)The fold destabilized compared to the noncanonical U•A-T base triple was calculated as \( K_{D,EMSA} \) or \( K_{D,MST} \) (Z•A-T)/\( K_{D,EMSA} \) or \( K_{D,MST} \) (U•A-T).

\(^{\text{c}}\)The fold destabilized compared to the noncanonical C•G-C base triple was calculated as \( K_{D,EMSA} \) or \( K_{D,MST} \) (Z•G-C)/\( K_{D,EMSA} \) or \( K_{D,MST} \) (C•G-C).

\(^{\text{d}}\)Reported \( K_{D,EMSA} \) values for unmodified base triples were obtained from Kunkler et al. (2019).

Modified RNA destabilizes an R•D-D triple helix
also contain a methyl group at position 5′) or due to the differences in the ribose pucker. As for the modified noncanonical \( \text{Z}\text{G-C} \) base triples, stabilities relative to \( \text{C}\text{G-C} \) ranged from no significant change (Um•G-C, m\(^5\)U•G-C, \( \Psi\text{•G-C} \), and m\(^8\)A•G-C), moderately destabilized between ∼2.5- to 3.5-fold (Am•G-C, I•G-C, and Gm•G-C), and destabilized greater than fivefold (m\(^5\)U•G-C and s\(^5\)U•G-C).

To the best of our knowledge, there are no published reports examining the stability of RNA modifications within noncanonical R•D-D base triples. Like the previously studied noncanonical R•D-D base triples, unmodified Z•G-C base triples can tolerate a wide range of bases while Z•A-T base triples are more sensitive to nucleotide mismatches (Table 1; Fig. 1D,H; Kunkler et al. 2019).

Overall, our EMSA results show that the modified base triples either have no effect or destabilize an R•D-D triple helix relative to the canonical U•A-T and C•G-C base triples. It should be noted that the relative stability of the base triples likely changes under different buffer conditions and sequence contexts, including the length of the triple helix, the ratio of U•A-T to C•G-C base triples, and nearest-neighbor effects (Leitner and Weisz 2000; Leitner et al. 2000; Plum 2002). For example, it is known that C•G-C-containing triple helices are stabilized when the pH is lowered (e.g., pH 5) due to the protonation of C•G-C base triples to C\(^+\)•G-C, leading to an additional hydrogen bond in each C•G interaction (Felsenfeld and Rich 1957; Felsenfeld et al. 1957; Völker and Klump 1994; Plum and Breslauer 1995; Leitner et al. 1998, 2000; Plum 2002). For D•D-D triple helices, the pK\(_a\) of C•G-C base triples flanked by T•A-T base triples is higher than neighboring C•G-C base triples or terminal C•G-C base triples (Leitner et al. 2000). In fact, the measured pK\(_a\) value of C•G-C flanked by T•A-T base triples is 7.4, suggesting that the C•G-C base triples in our R•D-D construct are possibly protonated at pH 7 (Leitner et al. 2000). Therefore, it is still necessary to examine the formation of each predicted triple helix and the effect on triple helix formation in the presence and absence of an RNA modification.

Regarding methodologies to measure K\(_D\) values, one major pitfall with using EMSAs to measure binding affinity is that, because EMSAs are a separation method, the experiment is no longer in equilibrium. Therefore, all measured K\(_{D,EMSA}\) values are apparent dissociation constants (K\(_{D,app}\)) and not true dissociation constants (K\(_D\)). To test if K\(_{D,EMSA}\) values accurately reflect true K\(_D\) values, we employed MST as a secondary method to measure K\(_D\) values.

**Measured binding affinities for our R•D-D triple helix are similar using EMSA and MST**

Though the phenomenon of molecules moving across temperature gradients was described over 150 years ago, using this principle to measure binding affinities has been made possible via microscale thermophoresis (MST) (Ludwig 1856; Seidel et al. 2013). MST is unique from other binding methods because it does not require immobilization, it does not depend on large size changes between apo and complex states, it can be used with complex buffers, and it uses small sample volumes (Seidel et al. 2013). To perform MST, one binding partner is fluorescently labeled (herein, Cy5-labeled D-D) and mixed with the other unlabeled binding partner at varying concentrations (herein, RNA from ∼2–50,000 nM). After equilibrium is reached, samples are loaded into capillaries and inserted into an MST instrument. The fluorescence of the sample is monitored at the same location that an IR laser heats the sample. Due to the phenomenon of thermophoresis, the molecules move either toward or away from the laser, thus changing the local fluorescence over time, which is recorded by the MST instrument. Because molecules move differently through a temperature gradient depending on size, charge, hydration shell, and conformation, the fluorescence of labeled binding partner moves differently between its bound and unbound states. Therefore, MST can measure true equilibrium dissociation constants (K\(_{D,MST}\)) for a variety of interactions, including protein–protein, protein–nucleic acid, nucleic acid–nucleic acid, protein–small molecule, and nucleic acid–small molecule interactions, including monitoring the formation of an R•D-D pyrimidine-motif triple helix (Seidel et al. 2013; Maldonado et al. 2018). Using the same R•D-D construct (except 5′-end has Cy5-fluorophore rather than a \([^{32}\text{P}]\)-radiolabel) and buffer conditions used for EMSAs, we used MST to measure the relative stabilities of the same Z•A-T and Z•G-C base triples (where Z = C, Cm, m\(^5\)C, U, Um, m\(^5\)U, m\(^3\)U, \( \Psi\), s\(^4\)U, A, Am, m\(^6\)A, I, G, Gm) (Supplemental Fig. S1).

First, we examined the Z•A-T base triples using an R•D-D construct with a 5′-Cy5-labeled purine-rich DNA strand (Fig. 2A). A change in relative fluorescence was observed with increasing concentrations of the RNA strand, presumably binding to dsDNA to form an R•D-D triple helix (Fig. 2B). K\(_{D,MST}\) values were determined for each of the four unmodified and the 11 modified Z•A-T base triples (Table 1; Fig. 2C,D; Supplemental Fig. S2A; Supplemental Table S2). The K\(_{D,EMSA}\) and MST values measured for the U•A-T base triple are within twofold (187 ± 25 nM and 109 ± 10 nM, respectively), showing the measured K\(_D\) values agree for both methods (Table 1; Kunkler et al. 2019). Further, both EMSA and MST measured similar stabilities relative to the canonical U•A-T base triple (Table 1; Figs. 1D, 2D). Both methods showed that all unmodified noncanonical Z•A-T base triples, their corresponding modified Z•A-T base triples, and m\(^3\)U•A-T all completely disrupted triple helix formation (Table 1; Figs. 1D, 2D; Kunkler et al. 2019). Further, both methods show that the Um•A-T base triple has no effect on stability and the s\(^5\)U•A-T base triple destabilizes the triple helix to ∼0.4 the stability of the unmodified U•A-T base triple (Table 1; Figs. 1D, 2D). However, while EMSAs showed a mild destabilization of ∼2.5-fold for the m\(^3\)U•A-T
and \( \Psi \)A-T base triples compared to the canonical U\( \Psi \)A-T base triple, MST showed no significant difference in their relative stabilities (Table 1; Figs. 1D, 2D). Overall, the \( K_{D,\text{EMSA}} \) and MST values and the relative stability of each Z\( \Psi \)A-T base triple are comparable for both methods.

We also tested the Z\( \Psi \)G-C base triples using the same MST setup (Table 1; Fig. 2E–H; Supplemental Fig. S2B; Supplemental Table S2). Like the canonical U\( \Psi \)A-T base triple, the \( K_{D,\text{EMSA}} \) and MST values for the canonical C\( \Psi \)G-C base triple are within twofold: 165 ± 18 nM and 128 ± 9 nM, respectively (Table 1). Further, most of the modified Z\( \Psi \)G-C base triples had similar destabilization relative to C\( \Psi \)G-C in their corresponding method: Cm\( \Psi \)G-C, m\( 5 \)C\( \Psi \)G-C, U\( \Psi \)G-C, Um\( \Psi \)G-C, m\( 5 \)U\( \Psi \)G-C, s\( 4 \)U\( \Psi \)G-C, A\( \Psi \)G-C, Am\( \Psi \)G-C, I\( \Psi \)G-C, and Gm\( \Psi \)G-C (Table 1). Two base triples, m\( 5 \)U\( \Psi \)G-C and \( \Psi \)U\( \Psi \)G-C, showed binding using both assays.
but had relative stabilities that differed by approximately fourfold, while two different base triples, m$^\text{A}$•G-C and G•G-C, which had $K_{\text{D,EMSA}}$ values of 251 ± 22 nM and 359 ± 33 nM, respectively, had no observed binding using MST (Table 1; Fig. 2H). These differences in the $K_{\text{D,EMSA}}$ and $K_{\text{D,MST}}$ values are likely due to some of the challenges with using MST to monitor R•D-D triple helix formation, which are discussed below. Overall, with a few exceptions, measured $K_{D}$ values are comparable between the two methods.

There are multiple advantages of using MST over EMSAs to measure binding affinities, such as MST measuring an insolution $K_D$ and rapid data collection (~15 min/run). However, we encountered some challenges when using MST for our R•D-D triple helix experiments. First, the relative fluorescence ($F_{\text{r}}$/F$_0$) for both the unbound and the bound states is not predictable or consistent among base triples (see Fig. 2B,F). Therefore, two assumptions are made when analyzing the data: (1) a change in the relative fluorescence corresponds to the formation of a triple helix and not another structure (e.g., an RNA–DNA double helix via strand displacement) and (2) if the relative fluorescence reaches a plateau, then that plateau is assumed to be the fully bound complex. Additionally, there is not much change in the relative fluorescence of dsDNA to the R•D-D triple helix (~0.01 units at the 10–11 sec time point), which reduces the signal-to-noise ratio (see Fig. 2B,F) and often leads to large errors in measured $K_{D}$ values among replicates (see Table 1). Finally, we noticed that the measured $K_{D,\text{MST}}$ values were larger the longer the IR laser was on, suggesting the heating of the sample leads to the dissociation of the R•D-D complex to dsDNA and RNA. Therefore, though MST is a powerful technique to measure the $K_{D}$ of an RNA–dsDNA complex, we suggest using another method, such as an EMSA or UV thermal denaturation assay, to monitor the formation of a triple helix.

Using either EMSA or MST, all RNA modifications could form the triple helix at the Z•G-C base triple site, though some of the Z•A-T base triples completely disrupted the R•D-D triple helix tested herein. In general, all tested modifications either have no significant effect on the stability or destabilize a triple helix. However, it is possible that one of the other naturally occurring RNA modifications not tested herein could significantly stabilize the formation of a triple helix. Further, RNAs with multiple modifications, modifications within a different sequence context, or modifications at a different location in the triple helix (i.e., not a central location) may have varying effects on the stability of an R•D-D triple helix. Overall, many of the RNA modifications examined destabilize the R•D-D triple helix, showing that minor changes to the RNA can perturb R•D-D triple helix formation.

**Known RNA modification sites in a human IncRNA destabilize an R•D-D triple helix in vitro**

RNA modifications affect the stability of R•D-D triple helices based on the results of our model R•D-D construct presented herein. Therefore, we searched for potential biologically relevant examples of an R•D-D triple helix being regulated by a modification switch. First, using RMVar, we examined published, high-confidence transcriptome-mapped modification sites (m$^5$C, m$^5$U, Ψ, I, and Nm) within a pyrimidine-rich segment of human IncRNAs that are at least 19-nt long, as our previous study showed a minimum of 19 base triples are required for triple helix formation in vitro (Supplemental File S1; Kunkler et al. 2019; Luo et al. 2021). Please note that m$^\text{A}$ was not included in our search as high-confidence sites are within the DRACH-motif (where D = A/G/U, R = A•G, H = A/C/U) and are therefore not within pyrimidine-rich sequences (Wei and Moss 1977; Narayan and Rottman 1988; Csepany et al. 1990; Harper et al. 1990; Dominissini et al. 2012; Meyer et al. 2012). Three IncRNAs fulfilled the aforementioned criteria: AC068025.2 (TAOK1), AL157886.1, and LINCO0940 (Fig. 3A). Further, Triplexator predicts that the RNA segment of both AL157886.1 and LINCO0940 may form triple helices within promoter regions (Supplemental File S2; Buske et al. 2012). Therefore, we performed EMSAs to monitor the formation of the triple helix using only the region-of-interest, for each IncRNA and DNA contain only canonical U•A-T and C•G-C base triples (Fig. 3; Supplemental Table S3). Of the three IncRNA segments tested, only the AL157886.1 segment showed triple helix formation ($K_{D,\text{EMSA}} = 131 ± 4$ nM), demonstrating the importance of experimentally testing predicted triple helices (Fig. 3). The lack of triple helix formation for the AC068025.2 and LINCO0940 segments is perhaps not too surprising, as previous studies show that pyrimidine-motif triple helices are unstable without C•G-C base triples (as in AC068025.2), with a high ratio of C•G-C to U•A-T (as in LINCO0940), and when neighboring base triples are C•G-C base triples (as in LINCO0940) (Roberts and Crothers 1996; James et al. 2003; Maldonado et al. 2018).

Because the unmodified AL157886.1 segment showed triple helix formation, we examined the stability of the AL157886.1•dsDNA triple helix with one m$^5$C modification at each of the three biologically relevant locations (C5, C15, and C17) and with m$^5$C at all three of the biologically relevant locations (Fig. 4). Compared to the unmodified AL157886.1 segment, the three single m$^5$C-modified AL157886.1 segments showed an average relative destabilization to only ∼0.6, while the triple m$^5$C-modified AL157886.1 segment destabilized to ∼−0.5 (Fig. 4). Though the AL157886.1•dsDNA triple helix was only modestly destabilized by the presence of m$^5$C modifications, this example gives precedence to the possibility of IncRNA•dsDNA triple helix formation being regulated by RNA modifications in vivo. However, further studies are needed to confirm if the formation of AL157886.1•dsDNA or other R•D-D triple helices are regulated by an RNA modification in cells.

Beyond the observed formation of an R•D-D triple helix in vitro, there are some important considerations before suggesting a modification can alter the stability of a cellular
Modified RNA destabilizes an R•D-D triple helix. First, for any cellular R•D-D triple helix, it is essential that the RNA of interest is localized to the nucleus where it can interact with gDNA and that the binding sites in both the IncRNA and the gDNA are accessible. Unfortunately, the localization of most IncRNAs, including AL157886.1, is unknown (Mas-Ponte et al. 2017; Wen et al. 2018). For modified cellular R•D-D triple helices containing nuclear-localized RNAs, the enzyme responsible for the modification must also localize within the nucleus. For the 11 RNA modifications tested herein, the enzymes responsible for the m^5^C, m^6^A, and I modifications in IncRNAs are present within the nucleus, but the enzymes responsible for m^5^U and Nm in IncRNAs and their subcellular localization are not known (Savva et al. 2012; Liu et al. 2013; Ping et al. 2014; Aguilo et al. 2016; Zhao et al. 2018; Bohnsack et al. 2019). Other points of consideration for modified IncRNAs are the nuclear abundance and if there are differences in the ratio of modified to unmodified nucleotide at the site of interest within different cell types, disease states, or developmental states, though a recent study suggests that IncRNAs can overcome the limitation of low-expression through the formation of phase-separated bodies (Unfried and Ulltisky 2022). Also notable is that we examined R•D-D triple helices with a 19-nt cut-off due to their formation in vitro, although shorter triple helices may occur in vivo due to the cellular microenvironment, such as the presence of nucleosomes (Kunkler et al. 2019; Maldonado et al. 2019; Unfried and Ulltisky 2022). Beyond the other IncRNA modifications not tested herein (e.g., 1-methyladenosine and N^6^,2′-O-dimethyladenosine) that may lead to changes in the stability of R•D-D triple helices in vivo, DNA modifications may also alter the stability of R•D-D triple helices (Sood et al. 2019). 5-methylcytosine (5mC), the most prevalent DNA modification in humans, is likely to affect the stability of R•D-D triple helices in vitro. Though 5mC would not disrupt hydrogen-bonding interactions with the RNA third strand, 5mC has been shown to change the stability and structure of dsDNA in vitro, such as increased thermal stabil-
FIGURE 4. EMSA results for modified 22-nt sequence from IncRNA AL157886.1 binding to D-D. (A) Table showing the R•D-D constructs and measured $K_{D,\text{EMSA}}$ values for the AL157886.1 segment containing no modifications, a single m$^\text{C}$ modification at three different sites, and three m$^\text{C}$ modifications. Red nucleotides indicate the m$^\text{C}$ location in each RNA. The putative Watson–Crick and Hoogsteen interactions are represented by a solid line (|) and a dot (*), respectively. The asterisk (‘) denotes the location of the 5′-[32P]-radiolabel. The reported $K_{D,\text{EMSA}}$ values are average of at least three independent replicates and the associated standard deviation. (B) Bar plot showing the stability of each modified triple helix relative to the unmodified triple helix. The relative stability was calculated as $K_{D,\text{EMSA}}$(unmodified)/$K_{D,\text{EMSA}}$(modified). Error bars are the standard deviation of at least three independent experiments.

**Materials and Methods**

**Oligonucleotide preparation**

Oligonucleotides were chemically synthesized and purchased as follows: DNA and unmodified RNA from Sigma-Aldrich and modified RNA from Dharmacon/Horizon Discovery. Oligonucleotide sequences are shown in Figures 1A,E, 2A,E, 3, and 4A. For electrophoretic mobility shift assays, oligonucleotides were prepared as before (Kunkler et al. 2019). Briefly, the purine-rich DNA oligonucleotides were 5′-end radiolabeled using γ$^\text{[32P]}$-ATP (Perkin Elmer) and T4 PNK (New England Biolabs) per the manufacturer’s protocol. Unreacted γ$^\text{[32P]}$-ATP was removed using a G25 MicroSpin Column (GE Healthcare). For microscale thermophoresis, the chemically synthesized 5′-Cy5-labeled purine-rich DNA was purchased from Sigma-Aldrich.

**Electrophoretic mobility shift assays (EMSA)**

The procedure was performed as in our prior study of R•D-D triple helix stability (Kunkler et al. 2019). Briefly, 10 nM of a pyrimidine-rich 31-mer DNA oligonucleotide and a 10 nM of the complementary purine-rich 5′-[32P]-radiolabeled 31-mer DNA oligonucleotide were mixed in Binding Buffer (25 mM sodium cacodylate [pH 7], 125 mM NaCl, 2 mM MgCl$_2$, 10% glycerol, and 0.1 mg/mL yeast tRNA) before heating at 95°C for 2 min and snap-cooling on ice for 2 min to form dsDNA. The pyrimidine-rich 22-mer RNA oligonucleotide was added at increasing amounts (10–100,000 nM) and equilibrated at 4°C for 24 h. Samples were loaded onto a 12% native polyacrylamide gel (19:1 acrylamide:bisacrylamide, 40 mM Tris-acetate at pH 7.0, 1 mM EDTA, and 10 mM MgCl$_2$) and resolved at 195 V for ~6 h at 4°C. Gels were then wrapped in plastic wrap and exposed overnight to a phosphorimager screen. The following day, the screens were scanned using an Amersham Typhoon (GE Healthcare) and quantified using ImageQuant software (GE Healthcare). A plot of triple helix concentration versus the RNA concentration was fit to the Hill equation (Equation 1) using OriginPro 2021 Graphing Software (OriginLab Corporation):

$$[ts] = \frac{[ds][ss]}{K_{D,\text{EMSA}} + [ss]^n}.$$  (1)

In Equation 1, [ts] is the triple helix concentration, [ds] is the initial Watson–Crick dsDNA concentration, [ss] is the initial RNA concentration, $K_{D,\text{EMSA}}$ is the apparent equilibrium dissociation constant, and $n$ is the degree of cooperativity. Here, all parameters ([ds], $K_{D,\text{EMSA}}$, n) were treated as variables. Extrapolated values for each base triple are in Table 1, Supplemental Tables S1 and S3, and Figure 4A.

**Microscale thermophoresis (MST)**

In Binding Buffer, 10 nM of a pyrimidine-rich 31-mer DNA oligonucleotide was mixed with 10 nM of its complementary purine-rich 5′-Cy5-labeled 31-mer DNA oligonucleotide. The solution was heated at 95°C for 2 min before snap-cooling on ice to form dsDNA. The pyrimidine-rich 22-mer RNA oligonucleotide was added at decreasing twofold serial dilutions from 51.2 µM to ~2 nM. After a 24-h equilibration at 4°C, the samples were analyzed on the NanoTemper Monolith nt. 155 Pico (NanoTemper Tech) at room temperature using Monolith standard capillaries at 5% excitation power and low MST power. Data were collected using MO.Control v1.6.1 and MO.Affinity Analysis v2.3 (NanoTemper Tech). We noticed that measured $K_{D,MST}$ values were larger than the longer the IR laser was powered on, likely due to temperature increase inducing dissociation of the R•D-D complex. Therefore, rather than allowing the MO.Affinity Analysis software to choose the time with the best signal-to-noise ratio,
In silico prediction of modified R•D-D triple helices

A list of human lncRNAs containing high-confidence modifications that we examined herein (m^5C, m^5U,Ψ, I, and Nm), and the modification within the lncRNA (see Supplemental File S1) was compiled on February 1, 2022 using RMVar (https://rmvar.renlab.org) (Luo et al. 2021). Please note that m^5U and s^4U have not been detected in human IncRNAs and that m^5A was not examined because the m^5A mark in human IncRNAs occurs primarily in the DRACH-motif, which has multiple purines that would destabilize pyrimidine-motif R•D-D triple helices (Wei and Moss 1977; Narayan and Rottman 1988; Cespany et al. 1990; Harper et al. 1990; Dominissini et al. 2012; Meyer et al. 2012; Kunkler et al. 2019). Next, RNA sequences surrounding the modification site were examined for pyrimidine-rich sequences with a high potential to form R•D-D triple helices. RNAs that had a modification within a pyrimidine-rich sequence at least 19-nt long, the previously reported shortest length of a stable pyrimidine motif R•D-D triple helix under our binding conditions, were tested for triple helix formation in their unmodified states using the same EMSA conditions as above (Kunkler et al. 2019). Further, Triplexator was used to predict if the RNA segments have the potential to bind within human promotor regions, as performed previously (Buske et al. 2012; Kunkler et al. 2019). Briefly, the RNA segments (see Fig. 3A) and all human promotor sequences (GRCh38, version Eldorado 04-2021; accessed March 7, 2022; www.genomatix.de) were inserted as the single- and double-stranded sequence, respectively, using Triplexator on its default settings for all parameters (Buske et al. 2012) (see Supplemental File S2). Please note that to compile Triplexator v.1.3.2 from source code successfully, the libboost_iosstreams-mt.so file integrated with Linux 3.10.0 or above was symlinked to an empty file named libboost_iosstreams-mt.so.5.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank two core facilities at the University of Notre Dame for access to the equipment used herein: the Biophysics Instrumentation Core (GE Healthcare Amersham Typhoon) and the Warren Family Research Center for Drug Discovery (NanoTemper Monolith nt. 155 Pico). We would also like to thank the fourth floor Stepan Chemistry Hall community for their helpful discussions and all members of the Brown Laboratory for their critical review of the manuscript. C.N.K. is a fellow of the Chemistry-Biochemistry-Biology Interface (CBBI) Training Program at the University of Notre Dame, supported by National Institutes of Health (NIH) grant T32GM075762, and a National Institute of General Medical Sciences National Research Service Award (NIGMS NRSA) Predoctoral Fellowship (F31GM136163). G.E.S. was funded by Notre Dame College of Science Summer Undergraduate Research Fellowship (COS-SURF). J.A.B. is funded by the National Institutes of Health (R00GM111430; R35GM133696), the Clare Boothe Luce Program of the Henry Luce Foundation, and start-up funds from the University of Notre Dame. Funding for open access charge was provided by NIH grant R35GM133696.

Received May 12, 2022; accepted June 29, 2022.

REFERENCES

Aguilo F, Li S, Balasubramaniam N, Sancho A, Benko S, Zhang F, Vashisth A, Rengasamy M, Andino B, Chen CH, et al. 2016. Deposition of 5-methylcytosine on enhancer RNAs enables the coactivator function of PGC-1α. Cell Rep 14: 479. doi:10.1016/j.celrep.2015.12.043

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-filpping mechanism. Nature 455: 818–821. doi:10.1038/nature07249

Amott S, Bond PJ. 1973. Structures for poly(U),poly(A),poly(U) triple stranded nucleonucleotides. Nat New Biol 244: 99–101. doi:10.1038/newb244099a0

Amott S, Selsing E. 1974. Structures of the polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT). J Mol Biol 88: 509–521. doi:10.1016/0022-2836(74)90498-7

Amott S, Bond PJ, Selsing E, Smith PJC. 1976. Models of triple-stranded polynucleotides with optimised stereochemistry. Nucleic Acids Res 3: 2470. doi:10.1093/nar/3.10.2459

Blank-Giwona J, Postepska-Igielska A, Grummt I. 2019. lncRNA KHPS1 activates a poised enhancer by triplex-dependent recruitment of epigenomic regulators. Cell Rep 26: 2904–2915. doi:10.1016/j.celrep.2019.02.059

Bohnsack KE, Höbartner C, Bohnsack MT. 2019. Eukaryotic 5-methylcytosine (m5C) RNA methyltransferases: mechanisms, cellular functions, and links to disease. Genes (Basel) 10: 120. doi:10.3390/genes10020102

Brown JA. 2020. Unraveling the structure and biological functions of RNA triple helices. Wiley Interdiscip Rev RNA 11: e1598. doi:10.1002/wrna.1598

Buske FA, Bauer DC, Mattick JS, Bailey TL. 2012. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res 22: 1372–1381. doi:10.1101/gr.130237.111

Cespany T, Lin A, Baldick CJ, Beemon K. 1990. Sequence specificity of mRNA N^6-adenosine methyltransferase. J Biol Chem 265: 20117–20122. doi:10.1016/S0021-9258(17)30477-5

de Bont R, van Larebeke N. 2004. Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis 19: 169–185. doi:10.1093/mutage/geh025

Debreuil YL, Expert-Bezançon C, Favre A. 1974. Structures for the polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT). Nucleic Acids Res 19: 3653–3660. doi:10.1093/nar/19.13.3653
changes in chromatin structure. Mol Cell 60: 626–636. doi:10.1016/j.molcel.2015.10.001
Povsic TJ, Dervan PB. 1989. Triple helix formation by oligonucleotides on DNA extended to the physiological pH range. J Am Chem Soc 111: 3059–3061. doi:10.1021/ja00190a047
Riley M, Maling B, Chamberlin MJ. 1966. Physical and chemical characterization of two- and three-stranded adenine-thymine and adenyne-uracil homopolymer complexes. J Mol Biol 20: 359–389. doi:10.1016/0022-2836(66)90069-6
Roberts RW, Crothers DM. 1996. Prediction of the stability of DNA triplexes. Proc Natl Acad Sci 93: 4320–4325. doi:10.1073/pnas.93.9.4320
Saintomé C, Clivio P, Favre A, Fourrey JL, Laugâa P. 1997. Site specific photo-crosslinking of single stranded oligonucleotides by a complementary sequence equipped with an internal photocleavable probe. Chem Commun 167–168. doi:10.1039/A606330F
Savva YA, Rieder LE, Reenan RA. 2012. The ADAR protein family. Genome Biol 13: 252. doi:10.1186/gb-2012-13-12-252
Seidel SAI, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Völker J, Klump HH. 1994. Electrostatic effects in DNA triple helices. Biochemistry 33: 13502–13508. doi:10.1021/bi00249a039
Wang S, Kool ET. 1995. Relative stabilities of triplex helices composed of combinations of DNA, RNA and 2′-O-Methyl-RNA backbones: chimeric circular oligonucleotides as probes. Nucleic Acids Res 23: 1157–1164. doi:10.1093/nar/23.7.1157
Wang S, Xu Y, Kool ET. 1997. Recognition of RNA by triplex formation: divergent effects of pyrimidine C-5 methylation. Bioorg Med Chem 5: 1043–1050. doi:10.1016/S0968-0896(97)00059-x
Wei CM, Moss B. 1977. Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry 16: 1672–1676. doi:10.1016/0005-2787(70)90156-5
Wen X, Gao L, Guo X, Li X, Huang X, Wang Y, Xu H, He R, Jia C, Liang F. 2018. IncSLdb: a resource for long non-coding RNA subcellular localization. Database 2018: 85. doi:10.1093/database/bay085
Xodo LE, Manzini G, Quadrifraglio F, van der Marel GA, van Boom JH. 1991. Effect of 5-methylcytosine on the stability of triple-stranded DNA—a thermodynamic study. Nucleic Acids Res 19: 5625–5631. doi:10.1093/nar/19.19.5625
Yang T, Low JJA, Woon ECY. 2020a. A general strategy exploiting m3C duplex-remodelling effect for selective detection of RNA and DNA m3C methyltransferase activity in cells. Nucleic Acids Res 48: e5. doi:10.1093/nar/gkz1047
Yang X, Liu M, Li M, Zhang S, Hiju H, Sun J, Mao Z, Zheng M, Feng B. 2020b. Epigenetic modifications of noncoding RNA: a novel dimension of cancer biology. Mol Cancer 19: 64. doi:10.1186/s12943-020-01159-9
Zhao Y, Dunker W, Yu YT, Karijolich J. 2018. The role of noncoding RNA pseudouridylation in nuclear gene expression events. Front Bioeng Biotechnol 6: 8. doi:10.3389/fbioe.2018.00008
Zhou Y, Kierzek E, Loo ZP, Antonio M, Yau YH, Chuah YW, Gerfman-Shoachat S, Kierzek R, Chen G. 2013. Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. Nucleic Acids Res 41: 6664–6673. doi:10.1093/nar/gkt352

MEET THE FIRST AUTHOR

Charlotte Kunkler

Meet the First Author is a new editorial feature within RNA, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of RNA and the RNA research community. Charlotte Kunkler is the first author of this paper, “A single natural RNA modification can destabilize a U+A+T-rich RNA:DNA:DNA triple helix.” Charlotte is a biochemistry graduate student in the Jessica A. Brown laboratory at the University of Notre Dame.

What are the major results described in your paper and how do they impact this branch of the field?

RNA modifications have previously been shown to alter the stability of RNA structures. Herein, we examined the stability of 11 dif-
ifferent naturally occurring RNA modifications at a single location within a U•A-T-rich pyrimidine-motif RNA•DNA-DNA triple helix. Our results indicate that, for the modifications tested, RNA modifications either destabilize or have no significant effect on the stability of the triple helix. Therefore, it is important to consider the modification status of an RNA predicted to form an RNA•DNA-DNA triple helix, as modifications can affect RNA•DNA-DNA triple helix stability.

What led you to study RNA or this aspect of RNA science?

The structure/function relationship of RNAs fascinates me. As a graduate student, I decided to study one particular RNA structure, the triple helix, to better understand the base triples that stabilize the triple-helical structure and how altering the formation of triple helices might lead to a change in function.

During the course of these experiments, were there any surprising results or particular difficulties that altered your thinking and subsequent focus?

I was surprised to find three different human biological examples that contain a modification within a pyrimidine-rich sequence. The original idea of searching for these sequences was a “shot in the dark,” but I am pleased that the search yielded fruitful results. I am excited to see if the examples we found or other RNA modifications within RNA•DNA-DNA triple helices have a functional change in cells.

What are some of the landmark moments that provoked your interest in science or your development as a scientist?

Growing up, I viewed science as any other school subject where a student learns facts about the subject from text written by experts. However, during my undergrad studies, I joined a research group where I was first introduced to using science as a tool for discovery. Science fascinates me because it is dynamic, with endless opportunity to learn and to contribute to the pool of knowledge. Giving back to my community is important to me, and science allows me to share what I have learned at the bench through the publication of my findings.

If you were able to give one piece of advice to your younger self, what would that be?

It is okay to be uncomfortable in science. Failed experiments, unexpected results, and seemingly contradictory conclusions are not a reflection on your worth as a person or scientist. Rather, uncomfortable situations are nucleation sites for personal growth and scientific discoveries with even greater impact.