Stability of an RNA●DNA–DNA triple helix depends on base triplet composition and length of the RNA third strand

Charlotte N. Kunkler, Jacob P. Hulewicz, Sarah C. Hickman, Matthew C. Wang, Phillip J. McCown and Jessica A. Brown[*]

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA

ABSTRACT

Recent studies suggest noncoding RNAs interact with genomic DNA, forming an RNA●DNA–DNA triple helix that regulates gene expression. However, base triplet composition of pyrimidine motif RNA●DNA–DNA triple helices is not well understood beyond the canonical U●A–T and C●G–C base triplets. Using native gel-shift assays, the relative stability of 16 different base triplets at a single position, Z●X–Y (where Z = C, U, A, G and X–Y = A–T, G–C, T–A, C–G), in an RNA●DNA–DNA triple helix was determined. The canonical U●A–T and C●G–C base triplets were the most stable, while three non-canonical base triplets completely disrupted triple-helix formation. We further show that our RNA●DNA–DNA triple helix can tolerate up to two consecutive non-canonical A●G–C base triplets. Additionally, the RNA third strand must be at least 19 nucleotides to form an RNA●DNA–DNA triple helix but increasing the length to 27 nucleotides does not increase stability. The relative stability of 16 different base triplets in DNA●DNA–DNA and RNA●RNA–RNA triple helices was distinctly different from those in RNA●DNA–DNA triple helices, showing that base triplet stability depends on strand composition being DNA and/or RNA. Multiple factors influence the stability of triple helices, emphasizing the importance of experimentally validating formation of computationally predicted triple helices.

INTRODUCTION

Triple helices (or triplexes) have been known to form *in vitro* for over 60 years (1–4). One common type is the major-groove pyrimidine motif triple helix, which is primarily composed of canonical T●A–T, U●A–U and C●G–C base triplets (where '●' and '–' represent Hoogsteen and Watson–Crick interactions, respectively). Here,

a pyrimidine-rich third strand (DNA or RNA) binds in a parallel orientation along the major groove of the purine-rich strand in a Watson–Crick double-stranded (ds)DNA or dsRNA (5,6). Triple helices, which we define as three or more consecutive base triplets, have been found in several naturally occurring RNAs, such as telomerase (7), *S*-adenosylmethionine (SAM)-bound riboswitches (8,9), U2–U6 base-paired spliceosomal RNAs (10) and metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) (11). More recently, it has been proposed that long noncoding (lnc)RNAs and genomic (g)DNA interact to form RNA●DNA–DNA triple helices (R●D–D) (12). One example of a lncRNA forming a pyrimidine motif lncRNA●gDNA triple helix is Fendrr (Fetal-lethal non-coding developmental regulatory RNA) (13,14). The Fendrr●gDNA triple helix silences *Foxf1* (encodes forkhead box F1) and *Pitx2* (encodes paired-like homeodomain 2) genes by recruiting either polycomb repressive complex 2 or trithorax group/mixed lineage leukemia complex to the promoter region (13,14). lncRNAs are not the only RNA class proposed to form triple helices with gDNA, as microRNAs may also form triple helices (miRNA●gDNA) to regulate gene expression (15,16). More recently, it was shown that intronic β-globin RNA interacts with an upstream regulator element through R●D–D triple helix formation, displacing transcription factors and RNA polymerase II to downregulate β-globin gene expression (17). With at least 27 000 known lncRNA genes, ∼1800 miRNA genes and numerous intronic RNAs in humans, RNA●gDNA triple helices potentially represent a large population of triple helices inside human cells (18–22). It has recently been shown that RNA●gDNA triple helices may be stabilized by the nucleosome, suggesting cellular components may stabilize triple-helix formation *in vivo* (23). Although many RNA●gDNA triple helices may exist *in vivo*, the stability of non-canonical R●D–D base triplets *in vitro* has not been investigated; therefore, the diversity of R●D–D base triplets may be greater than U●A–T and C●G–C base triplets.

---

[*]To whom correspondence should be addressed. Tel: +1 574 631 6486; Email: jbrown33@nd.edu

Unlike R•D–D triple helices, the relative stability of 16 different base triplets has been quantitatively determined for DNA•DNA–DNA (D•D–D) and RNA•RNA–RNA (R•R–R) triple helices (24–26). Although different assays and different triple-helical constructs preclude a direct comparison of base triplet stabilities, the studies conclude that the canonical T•A–T, U•A–U and C•G–C base triplets are among the most stable base triplets (24–26). For D•D–D triple helices, none of the non-canonical base triplets were within two-fold stability of the canonical T•A–T base triplet using a quantitative affinity cleavage assay (24). UV thermal denaturation assays showed that all 16 D•D–D base triplets allow for triple helix formation, with melting temperatures ranging from 38 to 11°C (26). For R•R–R triple helices, several base triplets support formation within two-fold of the canonical U•A–U base triplet: U•C–G, C•U–A, U•U–A, C•G–C and C•A–U base triplets based on native gel-shift assays and C•G–C and U•G–C based on the intronless β-globin reporter assay (25). Because relative stabilities of non-canonical base triplets vary for D•D–D and R•R–R triple helices, we predict that R•D–D triple helices have a unique set of non-canonical base triplets that support triple-helix formation. This prediction is also supported by UV thermal denaturation and quantitative affinity cleavage titration results, which show that strand identity affects the stability of R•D–D, D•D–D and R•R–R triple helices when the sequence is held constant (27–30). Together, these results indicate that the stability profiles of D•D–D and R•R–R base triplets are likely different from the stability profile of R•D–D base triplets.

In this study, the relative stability of 16 different base triplets at neutral pH was measured using native gel-shift assays by varying a single base triplet in an R•D–D triple helix. We further investigated the effects of consecutive non-canonical base triplets and length of RNA third strand on the relative stability of an R•D–D triple helix. Finally, we tested the relative stabilities of 16 base triplets within a D•D–D and R•R–R triple helix to directly compare with our results from an R•D–D triple helix. Collectively, these results show that the canonical U•A–T, T•A–T, U•A–U and C•G–C base triplets are the most stable base triplets, 11 non-canonical base triplets support R•D–D triple-helix formation and the relative stability profile of an R•D–D triple helix is distinct from that of a D•D–D and an R•R–R triple helix. Comparing our results on triple-helical stability to those previously published shows that multiple variables influence the stability of triple helices.

MATERIALS AND METHODS

Oligonucleotide preparation

Chemically synthesized DNA and RNA oligonucleotides were purchased from Sigma-Aldrich (Woodlands, TX, USA) and their sequences are shown in Figures 1A, E, I, 2A and 3. All radiolabeled oligonucleotides were 5′-end radiolabeled using γ-[32P]ATP (MP Biomedicals) and T4 PNK (New England Biolabs) per manufacturer's protocol. Unreacted γ-[32P]ATP was removed using a G25 microspin column (GE Healthcare).

In silico prediction of R•D–D triple helix in human noncoding RNAs and promoter DNA

A list of noncoding RNAs (ncRNAs) in humans (NON-CODEv5_human.fa) and the corresponding ID conversion file (NONCODEv5_source) were obtained from the NONCODE website (www.noncode.org/download) (31). The grep command was used to find our 22-nt RNA sequence (5′-UUUUUCUUUUZUUUUCUUUCUU-3′) in human ncRNAs (see Supplementary Table S1). Triplexator was used to predict if our 22-nt RNA could bind to human promoter DNA (32). First, a list of all human promoter sequences in Fasta format was obtained from Genomatix using the Gene2Promoter program of the Genomatix Suite package (www.genomatix.com). Using Triplexator, our search used our RNA sequence (5′-UUUUUCUUUUCUUUUCUUUCUU-3′) as the single-stranded sequence, human promoter sequences obtained from Genomatix as the double-stranded sequences and default settings were used for all parameters (32) (see Supplementary Table S2).

Electrophoretic mobility shift assays

In binding buffer (25 mM sodium cacodylate at pH 7.0, 125 mM NaCl, 2 mM MgCl$_2$, 0.1 mg/ml tRNA and 10% glycerol), 10 nM of the pyrimidine-rich 31-mer oligonucleotide and 10 nM of its complementary purine-rich 5′-[32P]-radiolabeled 31-mer were heated at 95°C for 2 min and then snap-cooled on ice for 2 min to form dsDNA or dsRNA. Increasing amounts of the pyrimidine-rich 22-nucleotide third strand were added (5–50 000 nM) and allowed to equilibrate 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 electrophoresed with running buffer (40 mM Tris-acetate at pH 7.0, 1 mM EDTA and 10 mM MgCl$_2$) at 195 V for ∼6 h at 4°C. Gels were wrapped in plastic wrap and exposed to a phosphorimager screen overnight, scanned using an Amersham Typhoon (GE Healthcare) and quantified using ImageQuant software (GE Healthcare). A plot of triple-helix formation versus the concentration of the third strand was fit to the Hill equation (Equation 1) for the R•D–D and D•D–D triple helices and to a quadratic equation (Equation 2) for the R•R–R triple helix using OriginPro 2018b graphing software (OriginLab Corporation).

$$[ts] = [ds][ss]^n / (K_{D,app}{}^n + [ss]^n) \qquad (1)$$

$$[ts] = 0.5(K_{D,app} + [ds] + [ss])$$
$$-0.5\left((K_{D,app} + [ds] + [ss])^2 - 4[ds][ss]\right)^{0.5} \quad (2)$$

[ts] is the concentration of the triple helix, [ds] is the initial concentration of the Watson–Crick dsDNA or dsRNA, [ss] is the initial concentration of the third strand, $K_{D,app}$ is the apparent equilibrium dissociation constant and n is the degree of cooperativity. All productive binding reactions for R•D–D and D•D–D triple helices reached a plateau, and the maximum [ts] was typically >7.5 nM (or >75%) (see

Supplementary Tables S3–S5). All parameters ([ds], $K_{D,app}$, n) were treated as variables when fitting data to Equation (1). Although all R•R–R triple helices tested showed some triple-helix formation, most did not reach saturation due to their large $K_{D,app}$ values (>30 μM); therefore, [ds] was fixed at 10 nM while $K_{D,app}$ was treated as a variable when fitting the data to Equation (2). Please note that binding data could not always be fit to a simple 1:1 quadratic binding model; therefore, a cooperative binding model was used during data analysis. At present, we do not have evidence for multiple binding sites nor more than one binding event. Nonetheless, we currently cannot rule out more complex binding behavior. Thus, we define our extrapolated equilibrium dissociation constants as $K_{D,app}$.

Circular dichroism spectroscopy

Three different solutions of oligonucleotides (see Supplementary Figure S2A) were prepared in CD buffer (25 mM sodium cacodylate at pH 7.0, 125 mM NaCl and 2 mM $MgCl_2$): (i) 4 μM RNA strand (R), (ii) 4 μM dsDNA (D–D) and (iii) 2 μM dsDNA and 2 μM RNA strand (R•D–D). First, DNA strands were added to CD buffer, heated for 2 min and then snap-cooled on ice for 2 min to form dsDNA. Then RNA was added and allowed to equilibrate at 4°C for 24 h. CD spectra were obtained at 20°C on a J-815 CD Spectrometer (JASCO) with a 1 mm quartz spectrophotometer cell (Starna Cells, Inc.) using the following parameters: 200–320 nm wavelengths with continuous scanning mode, 0.5 nm data pitch, standard sensitivity (±200 mdeg), 4 s digital integration time, 2 nm bandwidth and 100 nm/min scan speed. Data were an average of 5 scans, and molar ellipticity (θ) was calculated using Equation (3).

$$\theta = \text{mdeg} \times M / (10 \times L \times C) \tag{3}$$

mdeg is the millidegree rotation measured by the CD spectrometer at a specific wavelength, M is the mean residual weight (i.e. the average molecular weight of nucleotide monophosphates in the oligonucleotides (324.06, 326.81 and 326.09 g/mol for R, D–D and R•D–D, respectively)), L is the path length (1 cm) and C is the molar concentration of nucleic acids in solution as individual nucleotide monophosphates (0.0285, 0.081 and 0.0548 mol/L for R, D–D and R•D–D, respectively). The spectrum of buffer alone was subtracted from each sample spectrum. Data were an average of 5 scans, plotted as molar ellipticity versus wavelength. To determine if a structural change occurred upon RNA third strand binding to dsDNA, a molar ellipticity difference ($\Delta\theta$) spectrum was generated using Equation (4).

$$\Delta\theta = \theta_{R•D-D} - 0.5 (\theta_R + \theta_{D-D}) \tag{4}$$

θ_R, θ_{D-D} and $\theta_{R•D-D}$ are the molar ellipticity spectra calculated for R, D–D and R•D–D, respectively.

Stoichiometry of each strand in R•D–D triple helix

EMSA was performed as described above except the concentration of each strand was 2 μM, and all three strands were 5′-[32P]-radiolabeled. Bands representing dsDNA and the R•D–D triple helix were excised from the gel and crushed into pieces. After adding G50 buffer (20 mM Tris pH 7.5, 300 mM sodium acetate pH 5.2, 2 mM EDTA), the gel pieces were subjected to one freeze-thaw step and nutated at room temperature for 30 min before moving to 4°C overnight. Extracted oligonucleotides were ethanol precipitated and resuspended in water. Samples, along with standards consisting of three radiolabeled strands combined at 125, 250, 500 and 1000 nM, were loaded onto a 20% denaturing polyacrylamide gel (19:1 acrylamide:bisacrylamide, 1× TBE, 8 M urea) and electrophoresed with running buffer (0.5× TBE) at 30 W for ~7 h. Gels were wrapped in plastic wrap, exposed to a phosphorimager screen, scanned using an Amersham Typhoon (GE Healthcare) and quantified using ImageQuant software (GE Healthcare). Concentrations of extracted oligonucleotides were extrapolated from a linear standard curve generated from the band intensities at four different concentrations (125, 250, 500, 1000 nM) of radiolabeled oligonucleotides.

RESULTS

R•D–D triple helices can include most non-canonical base triplets

A three-strand construct was designed to determine the relative stability of 16 different R•D–D base triplets at a single position. The construct is comprised of 31-base pair dsDNA and a 22-nucleotide (nt) RNA third strand that can form a pyrimidine motif triple helix containing eighteen U•A–T base triplets, three C•G–C base triplets and one varying base triplet denoted as Z•X–Y (where Z = C, U, A, G and X–Y = A–T, G–C, T–A, C–G) (Figure 1A). The C•G–C base triplets are distributed asymmetrically so that the RNA third strand binds in a single register parallel to the Hoogsteen edge of the purine-rich DNA strand. Furthermore, each C•G–C base triplet is in the interior of the triple helix and is flanked by U•A–T base triplets so that the protonated state of the Hoogsteen cytosine in a C•G–C base triplet is more favorable based on studies of D•D–D triple helices (33,34). The varying base triplet Z•X–Y is centrally located within the triple helix, is flanked by U•A–T base triplets and the eight nearest base triplets are the same on both sides. Investigating this construct has physiological significance because the 22-nt RNA (where Z = C, U, A, G) is found in 33 human lncRNAs (Supplementary Table S1) and Triplexator predicts 1322 putative binding sites in human DNA promoters (Supplementary Table S2) (32, www.genomatix.de). All assays were performed using a pseudo-physiological buffer: a neutral pH of 7, 125 mM sodium chloride and 2 mM magnesium chloride.

EMSA was performed for each of the 16 base triplets, whereby an apparent equilibrium dissociation constant ($K_{D,app}$) was measured between the [32P]-radiolabeled dsDNA and the RNA third strand (Figure 1B and Supplementary Figure S1). Increasing concentrations of RNA showed a mobility shift consistent with the formation of an R•D–D triple helix and nearly 100% of the RNA at high concentrations was bound to dsDNA (Figure 1B, Supplementary Figures S1 and S2 and Table S3). Furthermore, circular dichroism (CD) spectroscopy indicated that a solution of R•D–D at a 1:1:1 ratio is structurally different from single-stranded RNA and dsDNA based on the individual
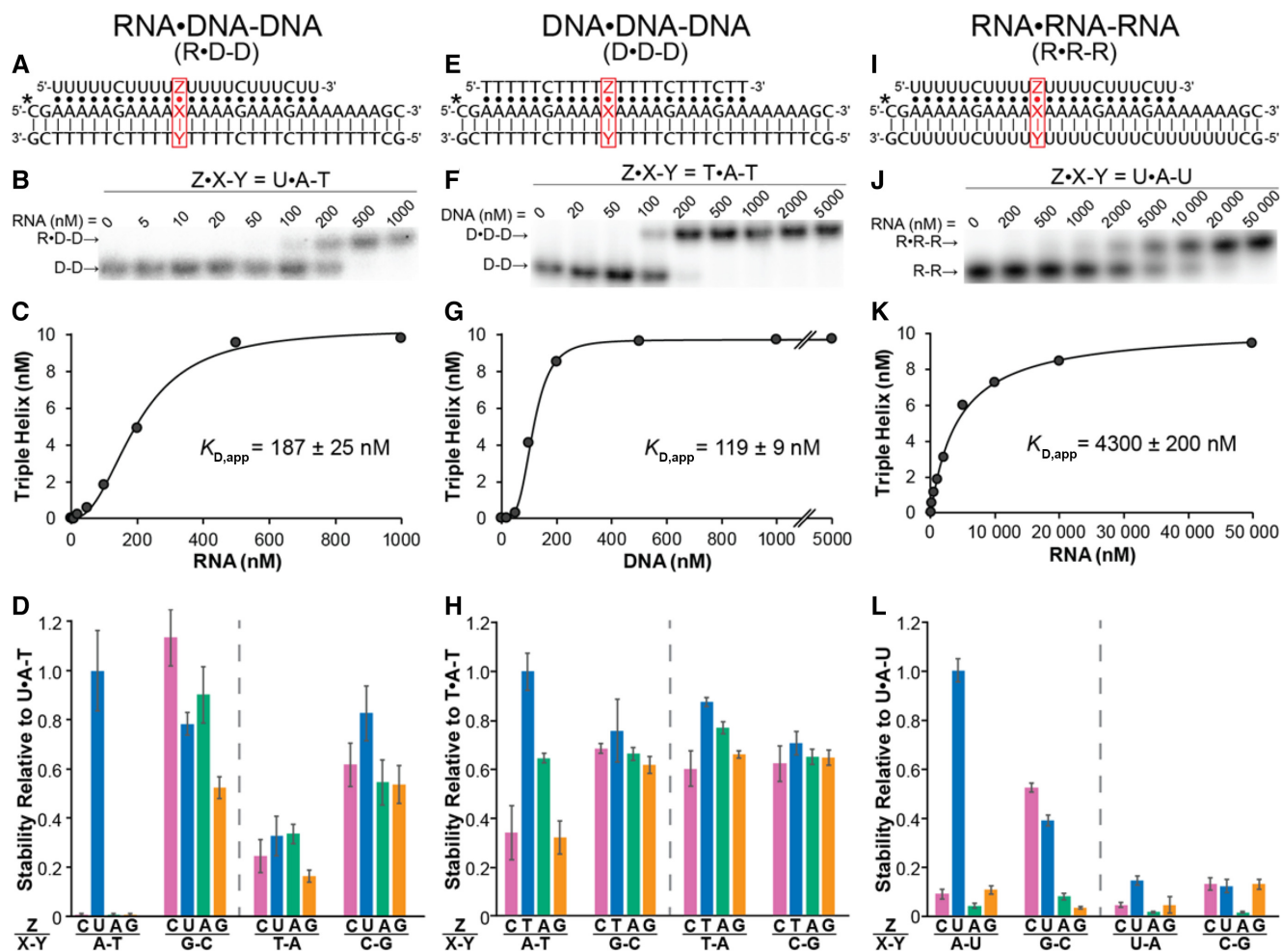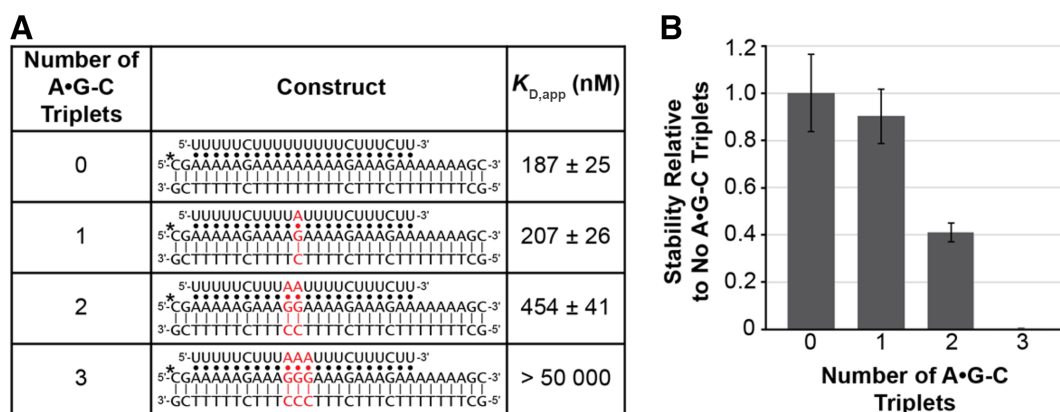
Figure 1. EMSA measurements of the relative stability of base triplets in R•D–D, D•D–D and R•R–R triple helices. Schematic is depicting (**A**) R•D–D, (**E**) D•D–D and (**I**) R•R–R triple helices with the varying position (Z•X–Y) having one of 16 possible base triplet combinations shown in red. The Watson–Crick and Hoogsteen interactions are represented by a solid line (|) and dot (•), respectively. The asterisks (*) denote the 5′-[^{32}P]-radiolabeled strand. Representative gel images are shown for the (**B**) U•A–T, (**F**) T•A–T and (**J**) U•A–U base triplets. Increasing amounts of the third strand were incubated with dsDNA (D–D) or dsRNA (R–R) for 24 h. The dsDNA/dsRNA were resolved from the triple helix (R•D–D, D•D–D and R•R–R) using native polyacrylamide gel electrophoresis and quantified by autoradiography. Representative binding curves are shown for (**C**) U•A–T, (**G**) T•A–T and (**K**) U•A–U base triplets. (**D**, **H**, **L**) The relative stability of each Z•X–Y base triplet is shown in the bar plots. Relative stability was calculated as (**D**) $K_{D,app}$(U•A–T)/$K_{D,app}$(Z•X–Y) for R•D–D base triplets, (**H**) $K_{D,app}$(T•A–T)/$K_{D,app}$(Z•X–Y) for D•D–D base triplets and (**L**) $K_{D,app}$(U•A–U)/$K_{D,app}$(Z•X–Y) for R•R–R base triplets. Each bar color represents a different nucleotide identity in the third strand (Z): pink for C, blue for U/T, green for A and orange for G. The dashed gray line separates the perfect polypurine-polypyrimidine Watson–Crick strands (left) from those with a single pyrimidine-purine interruption (right). Reported $K_{D,app}$ values are an average of at least three independent experiments and error bars represent standard deviation.
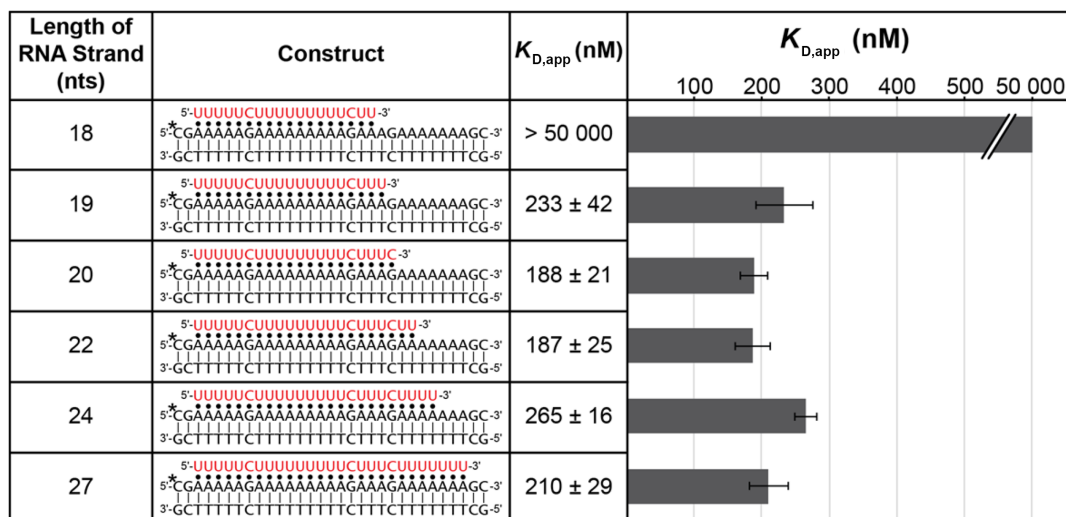
spectra and a CD difference spectrum showing significant peaks at ∼210, ∼255 and ∼270 nm (Supplementary Figure S3), which indicates the R•D–D mixture contains a structure distinctly different from single-stranded RNA and dsDNA and is similar to a previously studied D•D–D triple helix at pH 7 (35). Interestingly, the interaction between ds-DNA and RNA using EMSA showed positive cooperativity (n = 1.3–3.5) for all base triplets (Supplementary Table S3). This cooperativity has been observed previously (36); therefore, we quantitatively examined the stoichiometry of each strand in EMSA bands representing the dsDNA and R•D–D species. As expected, this analysis revealed a ratio of (1.1 ± 0.1):1 for purine-rich DNA : pyrimidine-rich DNA in dsDNA and (1.3 ± 0.2):(1.1 ± 0.1):1 for RNA : purine-rich DNA : pyrimidine-rich DNA in an R•D–D triple he-

lix (Supplementary Figure S4). Thus, the cooperative binding is not due to multiple RNA strands binding to dsDNA but instead likely reflects a conformational change in the dsDNA upon binding to RNA third strand or the zippering mechanism proposed for the formation of D–D double helices and D•D–D triple helices (6,37,38). These results established formation of an R•D–D triple helix with a binding stoichiometry of 1:1:1.

The $K_{D,app}$ values were determined for each of the 16 base triplets, and they ranged from 165 ± 18 nM for the canonical C•G–C base triplet to no binding in the presence of 5–50 000 nM RNA for C•A–T, A•A–T and G•A–T base triplets (Figure 1C, D and Table 1). All the $K_{D,app}$ values for R•D–D base triplets were normalized to the canonical U•A–T base triplet (Figure 1D and Table 1). As expected,

A

Number of A•G-C Triplets	Construct	$K_{D,app}$ (nM)
0	5'-UUUUUCUUUUUUUUUCUUUCUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	187 ± 25
1	5'-UUUUUCUUUUAUUUUCUUUCUU-3' 5'-CGAAAAAGAAAAGAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTCTTTCTTTCTTTTTTTCG-5'	207 ± 26
2	5'-UUUUUCUUUAAUUUUCUUUCUU-3' 5'-CGAAAAAGAAAGGAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTCCTTTTCTTTCTTTTTTTCG-5'	454 ± 41
3	5'-UUUUUCUUUAAAUUUUCUUUCUU-3' 5'-CGAAAAAGAAAGGGAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTCCCTTTCTTTCTTTTTTTCG-5'	> 50 000

B

Figure 2. EMSA measurements of the stability of R•D–D triple helices containing multiple consecutive non-canonical A•G–C base triplets. (**A**) Table showing the constructs and measured $K_{D,app}$ values for R•D–D triple helices with 0, 1, 2 and 3 consecutive A•G–C base triplets. Red nucleotides highlight the non-canonical A•G–C base triplets. The Watson–Crick and Hoogsteen interactions are represented by a solid line (|) and dot (•), respectively. The asterisks (*) denote the 5′-[³²P]-radiolabeled strand. Reported $K_{D,app}$ values are the average of at least three independent experiments ± the standard deviation. No binding was observed for three consecutive A•G–C base triplets when 5–50 000 nM of the RNA third stand was added. (**B**) Graph showing the relative stability of 0, 1, 2 and 3 consecutive non-canonical A•G–C base triplets. The relative stability was calculated as $K_{D,app}$(without A•G–C)/$K_{D,app}$(with A•G–C). Error bars are the standard deviation of at least three independent experiments.

Length of RNA Strand (nts)	Construct	$K_{D,app}$ (nM)	$K_{D,app}$ (nM)
18	5'-UUUUUCUUUUUUUUUCUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	> 50 000	
19	5'-UUUUUCUUUUUUUUUCUUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	233 ± 42	
20	5'-UUUUUCUUUUUUUUUCUUUC-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	188 ± 21	
22	5'-UUUUUCUUUUUUUUUCUUUCUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	187 ± 25	
24	5'-UUUUUCUUUUUUUUUCUUUCUUUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	265 ± 16	
27	5'-UUUUUCUUUUUUUUUCUUUCUUUUUUU-3' 5'-CGAAAAAGAAAAAAAAAGAAAGAAAAAAAGC-3' 3'-GCTTTTTCTTTTTTTTTCTTTCTTTTTTTCG-5'	210 ± 29	

Figure 3. EMSA measurements of the stability of R•D–D triple helices as the length of the third strand is varied. The constructs and measured $K_{D,app}$ values are shown for an R•D–D triple helix with 18-, 19-, 20-, 22-, 24- and 27-nt RNA third strand (red strand). The Watson–Crick and Hoogsteen interactions are represented by a solid line (|) and dot (•), respectively. The asterisks (*) denote the 5′-[³²P]-radiolabeled strand. Reported $K_{D,app}$ values are the average of at least three independent experiments ± standard deviation. The bar graph shows similar $K_{D,app}$ values for the 19- to 27-nt RNA third strands, while the 18-nt RNA third strand had no observable binding when 5–50 000 nM of the third strand was added. Error bars are the standard deviation of at least three independent experiments.

the canonical base triplets U•A–T and C•G–C exhibited the highest stability, although seven non-canonical base triplets exhibited a stability within two-fold of the canonical U•A–T base triplet (or $K_{D,app} \leq 374$ nM): A•G–C > U•C–G > U•G–C > C•C–G > A•C–G > G•C–G > G•G–C (Figure 1D and Table 1). These results suggested that the R•D–D triple helix can tolerate select non-canonical base triplets. Overall, the stability of our R•D–D triple helix appears to be controlled more by the identity of the Watson–Crick base pair (G–C ≈ C–G > T–A >> A–T) than by the nucleotide in the third strand (U > A ≈ C ≈ G) (Figure 1D). These results suggest that the nucleotide composition of R•D–D triple helices inside cells could include a combination of canonical and non-canonical base triplets.

Multiple non-canonical base triplets disrupt R•D–D triple helix formation

The relative stability of seven non-canonical R•D–D base triplets were within two-fold of the canonical U•A–T base triplet (Table 1). To test whether an R•D–D triple helix can form when there are multiple non-canonical base triplets, we replaced two and three consecutive U•A–T base triplets with A•G–C and performed native EMSA to determine $K_{D,app}$ values (Figure 2, Supplementary Figure S5 and Table S4). The stability of the R•D–D triple helix containing two consecutive A•G–C base triplets was only 2.2-fold weaker than one A•G–C base triplet (Figure 2). In contrast, no third strand binding to dsDNA was detected in the presence

Table 1. $K_{D,app}$ values for 16 different base triplets in R•D–D, D•D–D and R•R–R triple helices.

Base Triplet[a]	RNA•DNA-DNA		DNA•DNA-DNA			RNA•RNA-RNA		
	$K_{D,app}$ (nM)[b]	Normalized to U•A-T[c]	$K_{D,app}$ (nM)[b]	Normalized to T•A-T[c]	Fold ↑[d]	$K_{D,app}$ (nM)[b]	Normalized to U•A-U[c]	Fold ↓[e]
C•G-C	165 ± 18	1.1	174 ± 5	0.7	0.9	8200 ± 300	0.5	50
(T/U)•A-(T/U)	187 ± 25	1.0	119 ± 9	1.0	1.6	4300 ± 200	1.0	23
A•G-C	207 ± 26	0.9	179 ± 7	0.7	1.2	52 200 ± 8000	0.08	250
(T/U)•C-G	226 ± 32	0.8	168 ± 10	0.7	1.3	34 600 ± 8300	0.1	150
(T/U)•G-C	239 ± 14	0.8	157 ± 29	0.8	1.5	11 000 ± 600	0.4	46
C•C-G	303 ± 42	0.6	191 ± 22	0.6	1.6	32 200 ± 6000	0.1	110
A•C-G	343 ± 66	0.5	183 ± 10	0.7	1.9	249 400 ± 39 400	0.02	730
G•C-G	350 ± 52	0.5	184 ± 9	0.6	1.9	32 400 ± 5100	0.1	93
G•G-C	359 ± 33	0.5	192 ± 11	0.6	1.9	120 200 ± 16 200	0.04	340
A•(T/U)-A	559 ± 67	0.3	155 ± 5	0.8	3.6	223 300 ± 38 700	0.02	400
(T/U)•(T/U)-A	572 ± 159	0.3	136 ± 3	0.9	4.2	29 100 ± 4200	0.1	51
C•(T/U)-A	765 ± 195	0.2	197 ± 26	0.6	3.9	90 000 ± 22 500	0.05	120
G•(T/U)-A	1157 ± 154	0.2	180 ± 4	0.7	6.4	90 100 ± 46 400	0.05	78
A•A-(T/U)	> 50 000[f]	< 0.004	184 ± 5	0.6	> 270	98 400 ± 22 700	0.04	< 2
C•A-(T/U)	> 50 000[f]	< 0.004	349 ± 127	0.3	> 140	46 400 ± 9500	0.09	< 1
G•A-(T/U)	> 50 000[f]	< 0.004	368 ± 80	0.3	> 140	39 500 ± 6200	0.1	< 1

[a]Identity of varying base triplet Z•X-Y, where "•" and "-" represent Hoogsteen and Watson-Crick interactions, respectively. "(T/U)" represents DNA/RNA, respectively. Base triplets are listed from tightest to weakest binding in the R•D-D construct.
[b]Measured $K_{D,app}$ values reported as the average of at least three independent replicates ± the standard deviation. Here, green shading corresponds to a low $K_{D,app}$ value, yellow shading to an intermediate $K_{D,app}$ value, and red shading to a high $K_{D,app}$ value. Each heat map is relative to each triple helix.
[c]The relative stability compared to the canonical (T/U)•A-(T/U) base triplet for the given triple helix, which was calculated as $K_{D,app}$((T/U)•A-(T/U))/$K_{D,app}$(Z•X-Y).
[d]Fold tighter (↑) for D•D-D base triplet over R•D-D base triplet was calculated as $K_{D,app}$(Z•X-Y in R•D-D)/$K_{D,app}$(Z•X-Y in D•D-D).
[e]Fold weaker (↓) for R•R-R base triplet compared to R•D-D base triplet was calculated as $K_{D,app}$(Z•X-Y in R•R-R)/$K_{D,app}$(Z•X-Y in R•D-D).
[f]Showed no binding to dsDNA in the presence of 5-50 000 nM of the third strand RNA.

of 5–50 000 nM RNA when three consecutive A•G–C base triplets were present. Altogether, these results indicate that R•D–D triple helices can tolerate a limited number of consecutive non-canonical base triplets, although the number of consecutive non-canonical base triplets that an R•D–D triple helix can tolerate will likely vary depending on the identity of the base triplet and sequence context.

Formation of R•D–D triple helix depends on the length of the third strand

The stability of a triple helix may be affected by the length of the third strand. To test the dependence of an R•D–D triple helix on the length of the RNA third strand (i.e. the number of potential base triplets it can form), the relative stability of R•D–D triple helices was determined using RNA of six different lengths: 18, 19, 20, 22, 24 and 27 nts (Figure 3, Supplementary Figure S6 and Table S5). As expected, the RNA third strand exhibits a minimum-length requirement for binding to dsDNA. For our experimental system, the 18-nt RNA third strand does not bind whereas the 19- to 27-nt RNA third strands do bind. However, all the RNA third strands, from 19 to 27 nts, have approximately the same $K_{D,app}$ value with an average of ∼220 nM (Figure 3). These results show that the RNA third strand requires a minimum length to bind dsDNA, but increasing the length of the RNA third strand beyond the minimum length does not lead to a stronger interaction between the RNA third strand and dsDNA.

The relative stability of R•D–D base triplets is different from those of D•D–D and R•R–R base triplets

The relative stability of R•D–D base triplets presented herein showed a different trend than those observed previously for D•D–D and R•R–R base triplets (24,25). These differences could be due to different methodologies and/or different triple-helical constructs rather than the stability of base triplets. To make a direct comparison, the relative stabilities of 16 different base triplets in D•D–D (Figure 1E–H and Supplementary Figure S7) and R•R–R (Figure 1I–L and Supplementary Figure S8) triple helices were measured using our experimental setup. The constructs are identical to the R•D–D triple helix (Figure 1A) except all three strands were composed of DNA (Figure 1E) or RNA (Figure 1I), i.e. both the deoxyribose/ribose pentose and thymine/uracil bases were changed accordingly.

EMSA was performed for each of the 16 base triplets being varied at a single position Z•X–Y in both D•D–D and R•R–R triple helices (Figure 1E–L and Supplementary Figures S7 and S8). For the D•D–D base triplets, the $K_{D,app}$ values ranged from the canonical T•A–T base triplet at 119 ± 9 nM to the G•A–T base triplet at 368 ± 80 nM (Figure 1H, Table 1 and Supplementary Table S3). Thus, varying one base triplet in the D•D–D triple helix did not greatly alter the stability of this D•D–D triple helix, as only the base triplets C•A–T and G•A–T failed to bind within two-fold of the canonical T•A–T base triplet (i.e. $K_{D,app} \geq 238$ nM). The canonical C•G–C base triplet had similar stabil-

ity to the non-canonical base triplets (Figure 1H and Table 1). In contrast to the R•D–D triple helix, the stability of the D•D–D triple helix is controlled slightly more by the nucleotide of the third strand (T > A > C ≈ G) than the Watson–Crick base pair (T–A ≈ G–C ≈ C–G > A–T) (Figure 1D and H).

For the R•R–R base triplets, the $K_{D,app}$ values ranged from the canonical U•A–U base triplet at 4300 ± 200 nM to the A•C–G base triplet at 249 400 ± 39 400 nM (Figure 1L and Table 1). All $K_{D,app}$ values of R•R–R base triplets were normalized to the canonical U•A–U base triplet (Figure 1L and Table 1). The R•R–R triple helix is most stable when Z•X–Y is the canonical U•A–U base triplet. Only the canonical C•G–C base triplet was within two-fold of the canonical U•A–U base triplet (or have $K_{D,app} \leq 8600$ nM). Similar to the D•D–D triple helix, the stability of the R•R–R triple helix is controlled slightly more by the identity of the third strand base (U > C > G > A) than the Watson–Crick base pair (A–U ≈ G–C > C–G ≈ U–A) (Figure 1H and L).

Because the same sequence and methodology were used for all three triple helices, we can directly compare our findings for R•D–D, D•D–D and R•R–R base triplets (Table 1). Our results show that the stability of triple helices composed only of canonical base triplets (i.e. Z•X–Y = (T/U)•A-(T/U) or C•G–C) is as follows: D•D–D ≈ R•D–D > R•R–R, where R•R–R triple helices are ~35-fold weaker than R•D–D triple helices. As illustrated by the heat maps in Table 1, the relative stabilities of base triplets are unique to R•D–D, D•D–D and R•R–R triple helices. For example, the C•A–T and G•A–T base triplets are the least stable base triplets for the R•D–D and D•D–D triple helices but are moderately stable for the R•R–R construct (Table 1). Likewise, the A•C–G and A•U–A base triplets were the least stable base triplets for the R•R–R construct, moderately stable for the R•D–D construct and among the most stable for the D•D–D construct (Table 1). Our results show that the stabilities of base triplets in R•D–D, D•D–D and R•R–R triple helices are influenced by both strand identity and nucleotide composition.

DISCUSSION

Beginning in 2007, it was suggested that cellular ncRNAs might interact with gDNA via a triple helix as a mechanism to regulate gene expression (12,39). Since then, multiple examples of lncRNA•gDNA triple helices have been reported along with their mechanisms of controlling gene expression (12). Because the discovery and biological roles of lncRNA•gDNA triple helices are relatively recent, their base triplet composition and other factors that may influence the stability of these triple helices have not been investigated as extensively as D•D–D and R•R–R triple helices. By examining a U•A–T-rich R•D–D triple helix, our studies revealed the following: (i) all non-canonical base triplets except for C•A–T, A•A–T and G•A–T allow for triple-helix formation (Figure 1D and Table 1); (ii) the identity of the Watson–Crick base pair, X–Y, controls the stability of the base triplet more than the identity of the Hoogsteen base pair, Z•X (Figure 1D and Table 1); (iii) up to two consecutive A•G–C base triplets can be tolerated (Figure 2); and

(iv) the third strand has a sharp minimum length dependence (Figure 3), similar to previous studies on D•D–D triple helices (38). Increasing the length of the RNA third strand beyond the minimum length did not increase the stability of the triple helix (Figure 3), which may be due to the energetic penalty of a longer dsDNA segment undergoing a structural rearrangement upon binding a longer RNA third strand (40). Our study suggests that a large variety of R•D–D base triplets might exist inside cells considering that nine different non-canonical R•D–D base triplets support triple-helix formation (Figure 1D and Table 1) and that the thousands of lncRNAs, miRNAs and introns in humans could potentially interact with gDNA via a triple helix (18,19,22).

Currently, most lncRNA•gDNA interactions are initially identified using computational programs, such as Triplexator, TRIPLEXES or LongTarget (32,41–43). Triplexator and TRIPLEXES predict pyrimidine motif lncRNA•gDNA triple helices based on these default settings: (i) the canonical pyrimidine motif base triplets U•A–T and C•G–C observed previously (4) and (ii) forming at least 19 consecutive base triplets with at least 90% of the base triplets being the canonical U•A–T or C•G–C (32,41). LongTarget predicts pyrimidine motif R•D–D triple helices that can form at least 20 consecutive base triplets of U•A–T, C•G–C, U•G–C, G•G–C, C•C–G, U•C–G and G•T–A using the default settings (42,43). For a U•A–T-rich R•D–D triple helix, our results indicate that a minimum of 19–20 consecutive base triplets for predicted triple helices is reasonable (Figure 3), that <10% of the base triplets can be non-canonical base triplets except for C•A–T, A•A–T and G•A–T (Figures 1D, 2 and Table 1), that U•G–C, A•G–C, G•G–C, C•C–G, U•C–G, A•C–G and G•C–G (where underlined base triplets are already included in LongTarget) could also be considered by Triplexator, TRIPLEXES and LongTarget (Figure 1D and Table 1) and that non-canonical base triplet stability relies more on the Watson–Crick base pair (G–C ≈ C–G > T–A >> A–T) than the Hoogsteen base pair that currently dominates software algorithms (Figure 1D and Table 1) (32,41–43). Our results are for one R•D–D sequence context; therefore, more ncRNA•gDNA contexts (e.g. Fendrr, β-globin and PARTICLE) will need to be examined to fully understand how the energetics of triple helix formation depends on, for example, sequence composition, base triplet identity, length of triple helix and pH (13,14,17,44). Because multiple factors affect the stability of an R•D–D triple helix, it remains essential to validate the R•D–D triple helices identified using computational approaches. Validation studies performed in a test tube should also consider other factors, such as the intracellular concentration of lncRNA, the subcellular localization of the lncRNA, binding site accessibility of both the lncRNA and gDNA and cellular components, such as nucleosomes, that may stabilize the triple helix by creating unique microenvironments (23). Cell-based assays recently developed to isolate DNA-associated RNA and RNA-associated DNA provide an orthogonal approach to identify lncRNA•gDNA triple helices inside cells (45).

Our study showed that the relative stabilities of base triplets in R•D–D, D•D–D and R•R–R triple helices are

unique (Figure 1 and Table 1). Importantly, the relative stabilities of non-canonical base triplets in D•D–D and R•R–R triple helices that were measured previously exhibit a few similarities but, in general, show little overlap with those determined herein (Figure 1H, L and Table 1) (24,25). For example, when comparing our results to previous results for D•D–D triple helices using a quantitative affinity cleavage assay, the only base triplet to have approximately the same relative stability compared to T•A–T is the C•G–C base triplet (Figure 1H and Table 1) (24). Protonation at N3 of the Hoogsteen cytosine enables formation of two hydrogen bonds with G in the purine strand, making a protonated C•G–C base triplet more stable than a T•A–T base triplet (46). The C•G–C base triplet in our D•D–D triple helix likely exists in a predominantly deprotonated form at pH 7 because the C•G–C base triplet is 1.5-fold less stable than the T•A–T base triplet (Figure 1H and Table 1), and the pK_a of internal C•G–C base triplets in intermolecular pyrimidine motif D•D–D triple helices has been measured to be 5.6–6.2 (47). Furthermore, our study shows that the stability of all base triplets is within four-fold of T•A–T (Figure 1H and Table 1), whereas quantitative affinity cleavage assays showed all non-canonical base triplets as 15- to 1500-fold less stable than T•A–T (24). The G•T–A base triplet is typically one of the most stable non-canonical D•D–D base triplets (24,26,40,48–50); however, our study showed G•T–A as being comparable to most non-canonical base triplets (Figure 1H and Table 1). The similar relative stability of our D•D–D base triplets is likely because our D•D–D triple helix is too stable for our binding assay to detect stability changes for only a single base triplet. For R•R–R base triplets, three base triplets (C•G–C, U•G–C, G•A–U) have approximately the same relative stability compared to U•A–U both in our EMSA results (Figure 1L and Table 1) and in previous EMSA results (25). However, the relative stability profile of our R•R–R base triplets (Figure 1L and Table 1) was more similar to a profile determined previously using an intron-less β-globin reporter assay, as six base triplets (C•G–C, U•G–C, U•A–U, G•A–U, C•A–U, G•U–A) have approximately the same relative stability compared to U•A–U (25). These varying stability profiles in both D•D–D and R•R–R triple helices may be due to a combination of factors that differed among the studies: triple-helix construct (i.e. molecularity, length of the triple helix, location of varying Z•X–Y base triplet, sequence composition, ratio of canonical T•A–T:C•G–C base triplets and distribution of pH-sensitive C•G–C base triplets), conditions of binding reaction (i.e. buffer species and pH, salt identity and concentrations, polyamines, solvent, molecular crowding agents, temperature and equilibration time) and type of assay used to measure relative stabilities (24–26,40,48–50). Similar study-to-study variations have been observed for the stability of D•D–D base triplets and base pairs in dsRNA, both of which vary greatly depending on neighboring base triplets or base pairs (46,51,52). These results reinforce the importance of experimentally testing each computationally predicted triple helix, given the number of factors that can influence the overall stability of a triple helix.

Besides variation in base triplet stability, we also noticed there is little consensus on the relative stability of R•D–

D, D•D–D and R•R–R pyrimidine motif triple helices in the literature: a UV thermal denaturation assay for a two-strand construct (Watson–Crick hairpin and third strand) showed R•D–D > R•R–R > D•D–D at pH 5.5 (27), an affinity cleavage assay for a three-strand construct showed D•D–D ≈ R•D–D > R•R–R at pH 7.0 (28), a UV thermal denaturation assay for a three-strand construct showed R•D–D > D•D–D > R•R–R at pH 6.0 (29) and a UV thermal denaturation assay for a two-strand construct (circularized pyrimidine-rich strand and purine-rich central strand) showed R•D–D > D•D–D > R•R–R at pH 5.5 and D•D–D ≈ R•D–D ≈ R•R–R at pH 7.0 (30). Our results (Table 1) most closely resemble those of an affinity cleavage assay for a three-strand construct at pH 7.0 (D•D–D ≈ R•D–D > R•R–R) (28). All these studies, including ours, observe that R•D–D triple helices are more stable than R•R–R triple helices, although the relative stability of D•D–D triple helices varies greatly among the studies, from being the most to the least stable (27–30). The inconsistencies among the stability of base triplets and triple helices represent a gap in our understanding of triple helices and present a grand challenge in accurately predicting triple helices using a computational approach.

In summary, the *in vitro* stability of a pyrimidine motif R•D–D triple helix depends on multiple factors, including base triplet composition, length of the triple helix, molecularity, binding conditions and assay employed. Together these results show that the base triplet composition for RNA•gDNA triple helices inside cells may be more diverse than what is currently appreciated.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Felsenfeld,G., Davies,D.R. and Rich,A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.*, **79**, 2023–2024.
2. Lipsett,M.N. (1964) Complex formation between polycytidylic acid and guanine oligonucleotides. *J. Biol. Chem.*, **239**, 1256–1260.

3. Riley,M., Maling,B. and Chamberlin,M.J. (1966) Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. *J. Mol. Biol.*, **20**, 359–389.

4. Morgan,A.R. and Wells,R.D. (1968) Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J. Mol. Biol.*, **37**, 63–80.

5. Arnott,S. and Bond,P.J. (1973) Structures for poly(U).poly(A).poly(U) triple stranded polynucleotides. *Nat. New Biol.*, **244**, 99–101.

6. Arnott,S. and Selsing,E. (1974) Structures for the polynucleotide complexes poly(dA) · poly(dT) and poly(dT) · poly(dA) · poly(dT). *J. Mol. Biol.*, **88**, 509–521.

7. Theimer,C.A., Blois,C.A. and Feigon,J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell*, **17**, 671–682.

8. Gilbert,S.D., Rambo,R.P., Van Tyne,D. and Batey,R.T. (2008) Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nat. Struct. Mol. Biol.*, **15**, 177–182.

9. Huang,L. and Lilley,D.M.J. (2018) Structure and ligand binding of the SAM-V riboswitch. *Nucleic Acids Res.*, **46**, 6869–6879.

10. Yan,C., Hang,J., Wan,R., Huang,M., Wong,C.C. and Shi,Y. (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science.*, **349**, 1182–1191.

11. Brown,J.A., Bulkley,D., Wang,J., Valenstein,M.L., Yario,T.A., Steitz,T.A. and Steitz,J.A. (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.*, **21**, 633–640.

12. Li,Y., Syed,J. and Sugiyama,H. (2016) RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.*, **23**, 1325–1333.

13. Grote,P. and Herrmann,B.G. (2013) The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol.*, **10**, 1579–1585.

14. Grote,P., Wittler,L., Hendrix,D., Koch,F., Währisch,S., Beisaw,A., Macura,K., Bläss,G., Kellis,M., Werber,M. *et al.* (2013) The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell*, **24**, 206–214.

15. Toscano-Garibay,J.D. and Aquino-Jarquin,G. (2014) Transcriptional regulation mechanism mediated by miRNA–DNA•DNA triplex structure stabilized by Argonaute. *Biochim Biophys Acta - Gene Regul Mech.*, **1839**, 1079–1083.

16. Wang,S., Ke,H., Zhang,H., Ma,Y., Ao,L., Zou,L., Yang,Q., Zhu,H., Nie,J., Wu,C. *et al.* (2018) LncRNA MIR100HG promotes cell proliferation in triple-negative breast cancer through triplex formation with p27 loci. *Cell Death Dis.*, **9**, 805.

17. Zhou,Z., Giles,K.E. and Felsenfeld,G. (2019) DNA·RNA triple helix formation can function as a cis -acting regulatory mechanism at the human β-globin locus. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 6130–6139.

18. Hon,C.-C., Ramilowski,J.A., Harshbarger,J., Bertin,N., Rackham,O.J., Gough,J., Denisenko,E., Schmeier,S., Poulsen,T.M., Severin,J. *et al.* (2017) An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*, **543**, 199–204.

19. de Rie,D., Abugessaisa,I., Alam,T., Arner,E., Arner,P., Ashoor,H., Åström,G., Babina,M., Bertin,N., Burroughs,A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.

20. Soibam,B. (2017) Super-lncRNAs: Identification of lncRNAs that target super-enhancers via RNA:DNA:DNA triplex formation. *RNA*, **23**, 1729–1742.

21. Pasquier,C., Agnel,S. and Robichon,A. (2017) The mapping of predicted triplex DNA:RNA in the *Drosophila* genome reveals a prominent location in development- and morphogenesis-related genes. *G3 Genes Genomes Genet.*, **7**, 2295–2304.

22. Derrien,T., Johnson,R., Bussotti,G., Tanzer,A., Djebali,S., Tilgner,H., Guernec,G., Martin,D., Merkel,A., Knowles,D.G. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, **22**, 1775–1789.

23. Maldonado,R., Schwartz,U., Silberhorn,E. and Längst,G. (2019) Nucleosomes stabilize ssRNA-dsDNA triple helices in human cells. *Mol. Cell*, **73**, 1243–1254.

24. Best,G.C. and Dervan,P.B. (1995) Energetics of formation of sixteen triple helical complexes which vary at a single position within a pyrimidine motif. *J. Am. Chem. Soc.*, **117**, 1187–1193.

25. Brown,J.A., Kinzig,C.G., DeGregorio,S.J. and Steitz,J.A. (2016) Hoogsteen-position pyrimidines promote the stability and function of the MALAT1 RNA triple helix. *RNA*, **22**, 743–749.

26. Mergny,J.L., Sun,J.S., Rougée,M., Montenay-Garestier,T., Barcelo,F., Chomilier,J., Hélène,C. *et al.* (1991) Sequence specificity in triple-helix formation: Experimental and theoretical studies of the effect of mismatches on triplex stability. *Biochemistry.*, **30**, 9791–9798.

27. Roberts,R.W. and Crothers,D.M. (1992) Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science*, **258**, 1463–1466.

28. Han,H. and Dervan,P.B. (1993) Sequence-specific recognition of double helical RNA and RNA.DNA by triple helix formation. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 3806–3810.

29. Escudé,C., François,J.C., Sun,J.S., Ott,G., Sprinzl,M., Garestier,T. and Hélène,C. (1993) Stability of triple helices containing RNA and DNA strands: Experimental and molecular modeling studies. *Nucleic Acids Res.*, **21**, 5547–5553.

30. Wang,S. and Kool,E.T. (1995) Relative stabilities of triple helices composed of combinations of DNA, RNA and 2′-*O*-methyl-RNA backbones: Chimeric circular oligonucleotides as probes. *Nucleic Acids Res.*, **23**, 1157–1164.

31. Fang,S., Zhang,L., Guo,J., Niu,Y., Wu,Y., Li,H., Zhao,L., Li,X., Teng,X. and Sun,X. (2018) NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308.

32. Buske,F.A., Bauer,D.C., Mattick,J.S. and Bailey,T.L. (2012) Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.*, **22**, 1372–1381.

33. Leitner,D., Schröder,W. and Weisz,K. (2000) Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry.*, **39**, 5886–5892.

34. Plum,G.E. and Breslauer,K.J. (1995) Thermodynamics of an intramolecular DNA triple helix: a calorimetric and spectroscopic study of the pH and salt dependence of thermally induced structural transitions. *J. Mol. Biol.*, **248**, 679–695.

35. Soto,A.M., Loo,J. and Marky,L.A. (2002) Energetic contributions for the formation of TAT/TAT, TAT/CGC+, and CGC-/CGC+ base triplet stacks. *J. Am. Chem. Soc.*, **124**, 14355–14363.

36. Maldonado,R., Filarsky,M., Grummt,I. and Längst,G. (2018) Purine– and pyrimidine–triple-helix-forming oligonucleotides recognize qualitatively different target sites at the ribosomal DNA locus. *RNA*, **24**, 371–380.

37. Pörschke,D. and Eigen,M. (1971) Co-operative non-enzymatic base recognition III. Kinetics of the helix—coil transition of the oligoribouridylic · oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.*, **62**, 361–381.

38. Alberti,P., Arimondo,P.B., Mergny,J.L., Garestier,T., Hélène,C. and Sun,J.S. (2002) A directional nucleation-zipping mechanism for triple helix formation. *Nucleic Acids Res.*, **30**, 5407–5415.

39. Martianov,I., Ramadass,A., Serra Barros,A., Chow,N. and Akoulitchev,A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature.*, **445**, 666–670.

40. Yoon,K., Hobbs,C.A., Koch,J., Sardaro,M., Kutny,R. and Weis,A.L. (1992) Elucidation of the sequence-specific third-strand recognition of four Watson–Crick base pairs in a pyrimidine triple-helix motif: T.AT, C.GC, T.CG, and G.TA. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 3840–3844.

41. Kuo,C.C., Hänzelmann,S., Sentürk Cetin,N., Frank,S., Zajzon,B., Derks,J.P., Akhade,V.S., Ahuja,G., Kanduri,C., Grummt,I. *et al.* (2019) Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.*, **47**, e32.

42. He,S., Zhang,H., Liu,H. and Zhu,H. (2015) LongTarget: A tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. *Bioinformatics.*, **31**, 178–186.

43. Lin,J., Wen,Y., He,S., Yang,X., Zhang,H. and Zhu,H. (2019) Pipelines for cross-species and genome-wide prediction of long noncoding RNA binding. *Nat. Protoc.*, **14**, 795–818.

44. O'Leary,V.B., Smida,J., Buske,F.A., Carrascosa,L.G., Azimzadeh,O., Maugg,D., Hain,S., Tapio,S., Heidenreich,W.F., Kerr,J. *et al.* (2017) PARTICLE triplexes cluster in the tumor suppressor WWOX and may extend throughout the human genome. *Sci. Rep.*, **7**, 7163.

45. Sentürk Cetin,N., Kuo,C.C., Ribarska,T., Li,R., Costa,I.G. and Grummt,I. (2019) Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res.*, **47**, 2306–2321.

46. Roberts,R.W. and Crothers,D.M. (1996) Prediction of the stability of DNA triplexes. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 4320–4325.

47. Callahan,D.E., Trapane,T.L., Miller,P.S., Ts'o,P.O.P. and Kan,L.S. (1991) Comparative circular dichroism and fluorescence studies of oligodeoxyribonucleotide and oligodeoxyribonucleoside methylphosphonate pyrimidine strands in duplex and triplex formation. *Biochemistry*, **30**, 1650–1655.

48. Chandler,S.P. and Fox,K.R. (1995) Extension of DNA triple helix formation to a neighbouring (AT)n site. *FEBS Lett.*, **360**, 21–25.

49. Griffin,C. and Dervan,P. (1989) Recognition of thymine•adenine base pairs by guanine in a pyrimidine triple helix motif. *Science*, **245**, 967–971.

50. Wang,Y., Rusling,D.A., Powers,V.E., Lack,O., Osborne,S.D., Fox,K.R. and Brown,T. (2005) Stable recognition of TA interruptions by triplex forming oligonucleotides containing a novel nucleoside. *Biochemistry*, **44**, 5884–5892.

51. Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9373–9377.

52. Spasic,A., Berger,K.D., Chen,J.L., Seetin,M.G., Turner,D.H. and Mathews,D.H. (2018) Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res.*, **46**, 4883–4892.