Molecular Therapy
Nucleic Acids
Original Article

Thioflavin T Monitoring of Guanine Quadruplex Formation in the *rs689*-Dependent *INS* Intron 1

Ana Lages,[1] Christopher G. Proud,[1,2,3] John W. Holloway,[1] and Igor Vorechovsky[1]

[1]University of Southampton, Faculty of Medicine, Southampton SO16 6YD, UK; [2]Lifelong Health and Hopwood Centre for Neurobiology, South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia; [3]School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

The human proinsulin gene (*INS*) contains a thymine-to-adenine variant (*rs689*) located in the 3′ splice site (3′ ss) recognition motif of the first intron. The adenine at *rs689* is strongly associated with type 1 diabetes. By weakening the polypyrimidine tract, the adenine allele reduces the efficiency of intron 1 splicing, which can be ameliorated by antisense oligonucleotides blocking a splicing silencer located upstream of the 3′ ss. The silencer is surrounded by guanine-rich tracts that may form guanine quadruplexes (G4s) and modulate the accessibility of the silencer. Here, we employed thioflavin T (ThT) to monitor G4 formation in synthetic DNAs and RNAs derived from *INS* intron 1. We show that the antisense target is surrounded by ThT-positive segments in each direction, with oligoribonucleotides exhibiting consistently higher fluorescence than their DNA counterparts. The signal was reduced for ThT-positive oligonucleotides that were extended into the silencer, indicating that flanking G4s have a potential to mask target accessibility. Real-time monitoring of ThT fluorescence during *INS* transcription *in vitro* revealed a negative correlation with *ex vivo* splicing activities of corresponding *INS* constructs. Together, these results provide a better characterization of antisense targets in *INS* primary transcripts for restorative strategies designed to improve the *INS* splicing defect associated with type 1 diabetes.

INTRODUCTION

Guanine quadruplexes (G4s) are noncanonical secondary structures formed by G-rich nucleic acids stabilized by the stacking interactions of G tetrads held together by Hoogsteen base pairing.[1–3] They are formed in sequences containing repetitive stretches of two or more contiguous Gs (G-tracts), typically interrupted with 1–7-nt-long segments that comprise intramolecular G4 loops.[1–3] Their topologies can be influenced by a number of factors, including the loop length and the type of metal coordinating ions, but the rules governing their formation are not completely understood.[1,4] RNA G4s are more stable than DNA G4 structures and form readily *in vitro*, but the extent to which they affect gene expression pathways *in vivo*, such as processing of messenger RNA precursors (pre-mRNAs) or translation has remained doubtful, particularly in higher eukaryotes that are protein-rich and display global RNA G4 unfolding.[5–8]

G-tracts in exons and introns have been shown to influence the efficiency of pre-mRNA splicing for many years.[9–14] These motifs play a particularly important role in the accurate removal of short vertebrate introns.[9,10] For example, each of the seven G-triplets present in the 179-nt-long *INS* intron 1 incrementally enhanced its splicing.[15] A systematic deletion analysis of this intron revealed a potent splicing silencer surrounded by G-tracts and located just over 100 nt upstream of the 3′ splice site (3′ ss).[15] Targeting the silencer with antisense oligonucleotides increases the splicing efficiency of introns with the adenine allele at an SNP *rs689* (also known as INS-27 or HphI$^{+/-}$) to the levels observed for the thymine allele.[16] Apart from the cluster of genes encoding the major histocompatibility complex (MHC), the adenine at *rs689* shows the strongest genome-wide allelic association with type 1 diabetes and is more prevalent in Caucasian than in African populations.[17] The polymorphism resides 6 nt upstream of the *INS* intron 1-exon 2 junction and alters a key vertebrate pre-mRNA splicing signal known as the polypyrimidine tract, leading to higher intron 1 retention levels in exogenous transcripts with the *rs689* adenine as compared to the *rs689* uridine.[14] The increased intron retention was found in several cell types, including those derived from pancreatic β cells, and was proposed to result from impaired interactions of the adenine allele with one or more uridine-binding splicing factors that recognize the polypyrimidine tract and are important for selection of intron-lariat branch points and the 3′ ss.[14,15] *rs689* may therefore represent a causal disease variant by reducing proinsulin expression from the adenine-containing chromosomes in the developing thymus. This reduction was proposed to result from a diminished intron-mediated enhancement of translation, from an upstream open reading frame within the retained intron that curtails canonical *INS* translation, from stable RNA secondary structures within the extended G-rich 5′ UTR of intron 1-containing transcripts that leave the nucleus, or from a combination of these factors.[14,15] Collectively, these studies show that antisense strategies could modulate human proinsulin expression in a *rs689*-dependent manner through the intronic splicing silencer surrounded by G4s. However, it is unclear how G4 formation affects function of the silencer in transcripts that retain intron 1.

Received 23 November 2018; accepted 27 April 2019;
https://doi.org/10.1016/j.omtn.2019.04.026.

Correspondence: Igor Vorechovsky, University of Southampton, Faculty of Medicine, MP808, Tremona Road, Southampton SO16 6YD, UK.
E-mail: igvo@soton.ac.uk

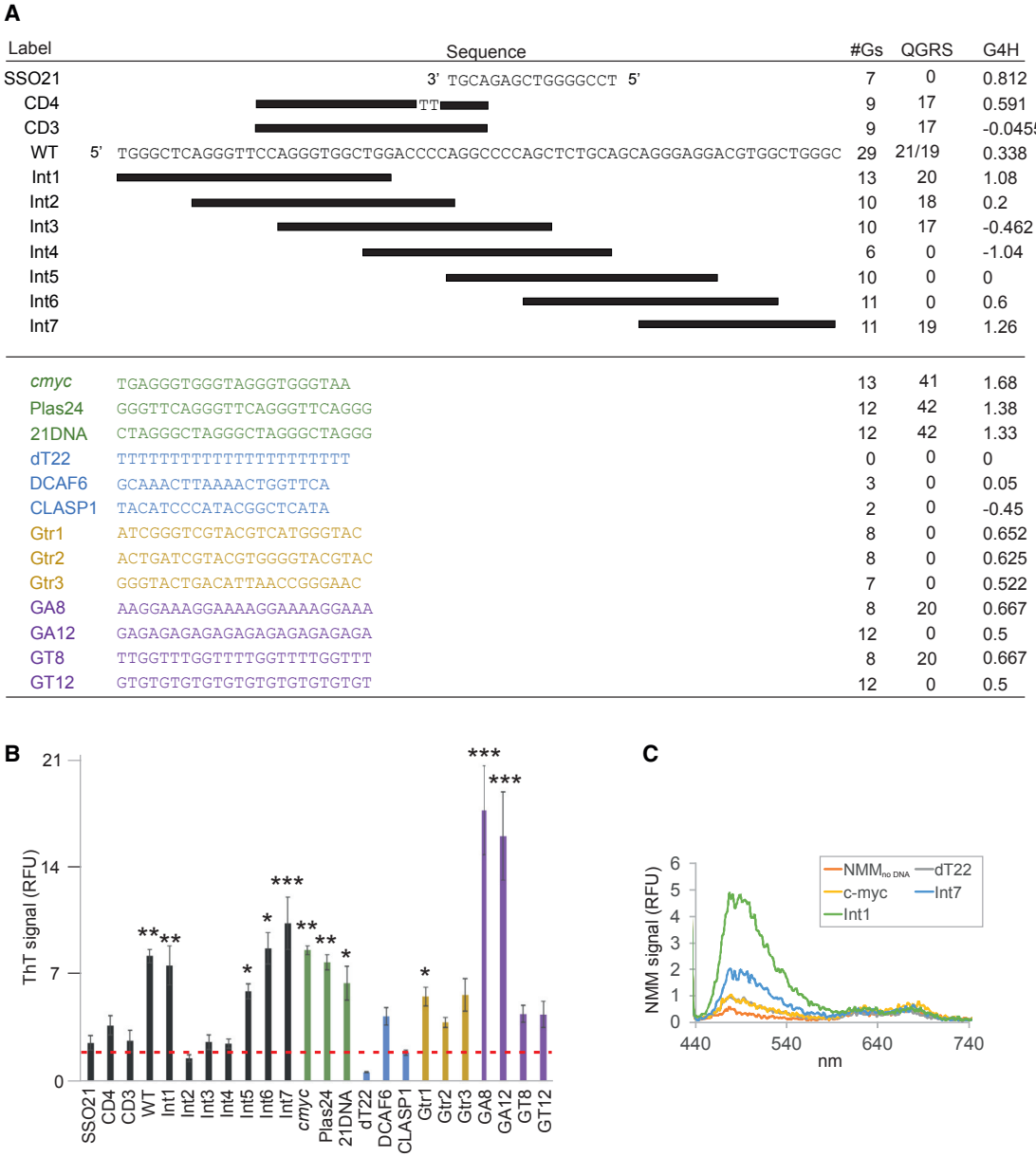

Figure 1. ThT Monitoring of *INS* Intron-1-Derived DNA Oligonucleotides

(A) Tested (upper panel) and control (lower panel) oligonucleotides. Each tested oligonucleotide is shown as a black rectangle representing the indicated portion of the wild-type (WT) partial sequence of *INS* intron 1 (GenBank: AH002844.2). SSO21, antisense oligonucleotide that reduced *INS* intron 1 retention *ex vivo*.[16] Int7 folds into parallel G4, as shown by far-UV circular dichroism and NMR.[16] CD3 forms a hairpin/G4 equilibrium *in vitro*; two C>T mutations (CD4) shifted the equilibrium toward G4 formation.[16] The number of Gs in each oligonucleotides, QGRS Mapper scores,[28] and G4 Hunter[29] scores are shown in the last three columns. (B) Mean relative fluorescence intensity (relative fluorescence units [RFU]) of 80 μM ThT at 508 nm in the presence of the indicated DNAs. Tested oligonucleotides are in black, positive G4 controls are in green, negative controls are in blue, and additional controls (see the main text) are in orange and purple. Error bars denote the SD of three independent ThT assays. *$p < 0.05$, **$p < 0.01$, or ***$p < 0.001$ for comparisons with negative controls (in blue). The mean for CLASP1 is denoted by a dashed red line as an arbitrary threshold. ThT-only controls are not shown. (C) NMM spectra for a subset of INS-derived DNA oligonucleotides and controls.

Detection of G4s has been facilitated recently by fluorescent dyes used as *in vitro* light-up structural probes (reviewed by Kwok and Merrick[3]). In particular, thioflavin T (ThT) provides a sensitive and selective means of detecting G4s and their topologies.[18–24] A selectivity of ThT for G4 structures has been demonstrated by large increments in the fluorescence emission at around 490 nm, which contrast to a low

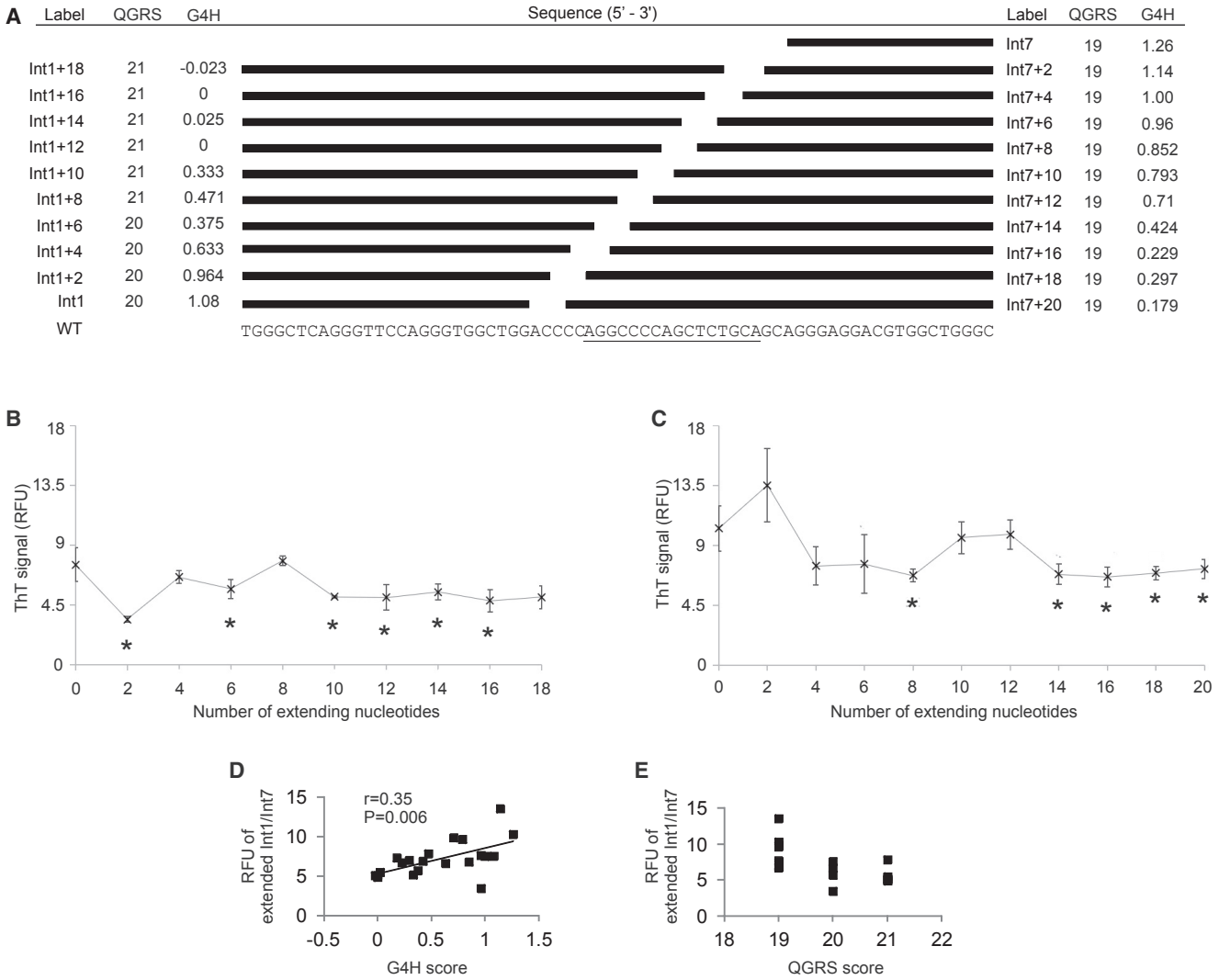

Figure 2. ThT Fluorescence of the *INS* Intron-1-Derived Int1 and -7 that Extended into the Antisense Target Region

(A) Schematics of tested DNA oligonucleotides (black rectangles) and their G scores.[28] The antisense target is underlined. (B and C) Mean fluorescence intensity of ThT at 508 nm in the presence of oligonucleotides extended from Int1 (B) and Int7 (C) into the target region. Error bars denote SD from three independent fluorescence assays. Asterisks denote significant p values < 0.05, when comparing fluorescence intensity of the extended nucleotide with the non-extended one. (D and E) Correlation between ThT fluorescence trace and G4H (D) or QGRS Mapper (E) scores.

signal in the presence of double- or single-stranded DNA sequences or water controls.[22] Guanine is the most favorable nucleobase for ThT binding, although ThT bound to a non-G-quadruplex structure may also yield an elevated fluorescence signal.[25,26] Nevertheless, the ThT probe can specifically recognize RNA G4s that adopt, in an exclusive manner, all-parallel conformations independent of their sequences and experimental conditions.[6,27] However, practical applications of ThT G4 monitoring have so far been underexplored.

In this study, we employed ThT to investigate the propensity of splicing regulatory motifs surrounding the antisense intron retention target to form G4s *in vitro*.

RESULTS

To evaluate G4 formation in previously identified splicing regulatory sequences of *INS* intron 1, we first tested a set of overlapping, short, synthetic nucleic acids by using a G4-sensitive fluorescent probe, ThT (Figure 1A). The region was centered on the optimal antisense target for reducing *INS* intron 1 retention that is surrounded by G-tracts.[16] Apart from tested *INS*-derived oligonucleotides, we examined G-rich and G-poor positive and negative controls, respectively, and an antisense oligonucleotide, SSO21, that reduced intron 1 retention in exogenous transcripts in several cell lines.[16] As additional negative controls, we used G-rich oligonucleotides (Gtr1-3 shown in orange and GA12 and GT12 shown in purple in Figure 1A) that were not predicted by

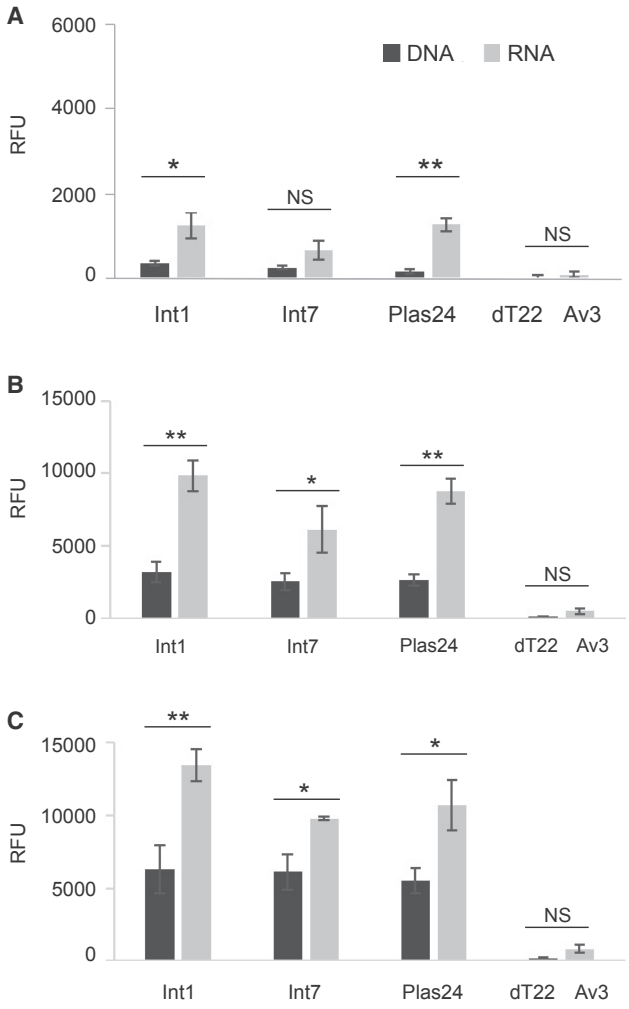

Figure 3. ThT G4 Screening in *INS* DNA and RNA Oligonucleotides

Oligoribonucleotide concentrations were 2 (A), 10 (B), and 20 (C) μM. AV3[15] and dT22 were used as negative controls. Error bars denote SDs from three independent assays. *p < 0.05 or **p < 0.01, determined by Tukey's multiple-comparisons test. NS, corrected p values > 0.05.

a scoring method[28] to fold into G4 structures. As positive controls, we employed their high-score counterparts (GA8 and GT8), as well as the widely used G4-forming c-*myc*, Plas24, and 21DNA (*DBF4B*).[21]

Figure 1B shows a summary of mean ThT fluorescence levels for 24 oligonucleotides tested. Repeated measurements revealed that G-runs, either upstream or downstream of the antisense intron retention target, which is delineated by SSO21, were prone to G4 formation *in vitro*. The correlation between fluorescence levels of all oligonucleotides and G4 prediction scores tended to be significant, with a lower correlation with the QGRS Mapper[28] than with the G4Hunter (G4H)[29] predictions ($r = 0.24$, $p = 0.13$ versus $r = 0.34$, $p = 0.05$, respectively). Upstream of the target, we observed strongly positive ThT values for Int1, but not for the overlapping Int2, which has four cytosines adjacent to the target. The highest fluorescence signal downstream of the target was found for Int7, which was higher than a longer wild-type (WT) control. Positive signals for Int1 and Int7 as compared to controls were also observed with N-methyl mesoporphyrin IX (NMM), which binds G4[30] (Figure 1C), but the highest signal was produced by Int1. With ThT, several negative controls generated fluorescence significantly above an arbitrary threshold of CLASP1, which contains only two Gs (Figure 1B). Adenine-rich negative controls (for example, DCAF6) had higher ThT values than adenine-poor negative controls, with the highest ThT fluorescence detected for GA-rich oligonucleotides GA8/GA12. We conclude that the DNA segments surrounding the antisense target for reducing *INS* intron 1 retention are prone to forming G4 structures *in vitro*.

To explore if G4 formation in these pre-mRNA segments could be influenced by the intron retention target, we examined ThT fluorescence for a set of oligonucleotides extending the highly positive Int1 and -7 into the flanking target by 2 nt increments (Figure 2A). The extended oligonucleotides were still prone to G4 formation; nevertheless, we observed an overall reduction in fluorescence as their length increased, with significant fluctuations in ThT signals between oligonucleotides ($p = 0.00002$ for the Int1 group and $p = 0.0006$ for the Int7 group, Friedman's test; Figures 2B and 2C). The reduction was consistent with the low ThT fluorescence of Int2, -3, and -4 and a lower signal for WT as compared to Int7 (Figure 1B). The fluorescence levels of extended oligonucleotides gave significant positive correlation with G4H scores but not with QGRS scores (Figures 2D and 2E). Together, these results show that the antisense target sequence has the potential to reduce formation of G4s, both upstream and downstream.

We then compared the ThT fluorescence of the DNA and RNA counterparts of selected *INS*-derived sequences and a control non-G4 pair. These measurements revealed higher fluorescence levels for oligoribonucleotides at several equimolar concentrations of nucleic acids, although not strictly statistically significant in each case (Figure 3). We also examined their mutated versions (Figure 4), which reduced (M3) or increased (M5 and M6) intron retention levels when the same mutations were introduced in the WT *INS* splicing reporter construct. In addition, we tested CD2, a 20-mer derived from a region upstream of the antisense target that formed stable hairpin structures rather than G4s by NMR spectroscopy and circular dichroism.[16] Mutations in M3 reduce hairpin formation while still maintaining some G4 potential, whereas M5 retains the hairpins but not G4.[16] Formation of both hairpins and G4s was reduced in M6,[16] consistent with a background ThT signal (Figure 4). The coexistence of significant hairpin structures and G4s previously detected in CD3 by NMR[16] did not translate into a high ThT signal. A significant fluorescence enhancement was observed for CD2, which contains four of five predicted Int1 G-tracts, but removal of a G-triplet from the 5′ end of Int1 in CD2 diminished the ThT fluorescence. Correlation between oligoribonucleotide fluorescence signals and QGRS ($r = 0.80$, $p = 0.004$) or G4H ($r = 0.78$, $p = 0.005$) scores was highly significant. Overall, these results point to the critical importance of this G-triplet for G4 formation upstream of the antisense target.

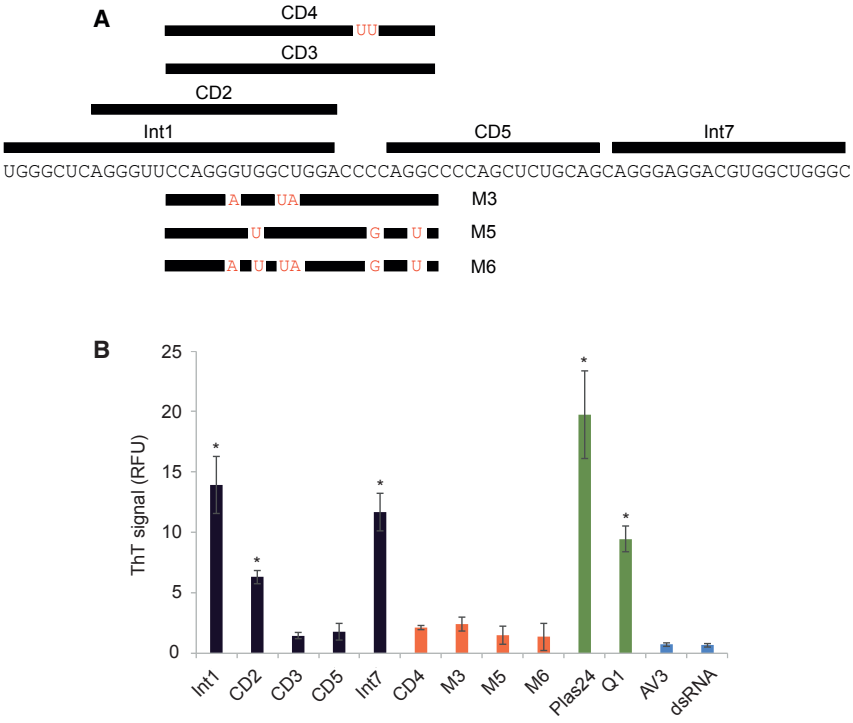

Figure 4. ThT Monitoring of Oligoribonucleotides Derived from the Antisense Target Region

(A) Schematics of tested oligoribonucleotides (horizontal black bars below and above the primary transcript). Mutations (in red) altered G4/hairpin equilibrium (CD4) and/or intron retention (M3, M5, and M6).[16] (B) Mean fluorescence intensity of 80 μM ThT at 508 nm in the presence of the WT oligoribonucleotides (black), their mutants (red), and positive (green) and negative (blue) controls. Error bars represent SDs from three independent fluorescence assays. Asterisks denote significant p values < 0.01.

Recently, Endoh and co-workers[31] described real-time monitoring of G4s, by using ThT. Applying this technique to *INS* intron 1, the fluorescence of the ThT signal was monitored during *in vitro* transcription of *INS* WT and mutated templates (Figure 5). The templates were prepared using proinsulin splicing reporter constructs described previously, which included mutations in G-runs and a deletion (del5) of the silencer region.[15,16] ThT fluorescence was measured during transcription of WT, del5, M3, and M5 constructs, and a control run-off transcript with several predicted G4s (Figures 5A–5F). In parallel to guanosine triphosphate (GTP) reactions, nascent transcripts were also synthesized in the presence of 7-deazaguanine (7-deazaG), which eliminates Hoogsteen interactions while still permitting canonical base pairing.[32] Finally, ThT fluorescence was related to the splicing activity of WT and mutated splicing reporters, as previously quantified by intron 1 retention levels following transient transcriptions into HEK293 cells.[15,16]

We observed a significant enhancement of ThT fluorescence during transcription reactions in the presence of GTP, as compared with the same transcripts synthesized in the presence of 7-deazaG or no-template controls (Figures 5A–5E). Importantly, ThT fluorescence in the transcripts tested tended to show inverse correlation with intron 1 retention levels at each time point (Figure 5F) and appeared to reflect the fraction of low-mobility conformers seen for the same RNAs on native gels (Figure 5G). The largest signal increase was observed for the del5 transcript (Figure 5B), which most improved splicing following a systematic deletion analysis of the whole intron.[15] This RNA also showed a reduced signal from the low-mobility conformer on a native gel (Figure 5G). Collectively, these results showed that during transcription *in vitro*, a splicing regulatory segment of *INS* intron 1 folds into ThT-positive structures and that the ThT signal from these transcripts tends to correlate inversely with their intron retention levels.

DISCUSSION

Understanding fully the role of human DNA variants in disease pathogenesis is a daunting challenge. A detailed elucidation of how downstream gene expression pathways are affected by each allele is often a prerequisite for developing efficient therapeutic interventions, as exemplified by the initial promise and success of antisense-based therapy for spinal muscular atrophy.[33] Our study focused on a common *INS* variant that leads to a minor splicing defect of intron 1, which extends the 5′ UTR in a fraction of mature transcripts.[14–16] The adenine allele is associated with genetic susceptibility to type 1 diabetes, and this association is the strongest among any non-MHC genes,[17] consistent with proinsulin acting as a critical autoantigen in the disease initiation.[15,34] By disrupting the polypyrimidine tract, adenine at *rs689* increases retention of this already weakly spliced intron in polyadenylated transcripts, extending the 5′ UTR and introducing translation-inhibitory sequences in a fraction of exported mRNAs.[14,15] Intron 1 retention can be ameliorated *ex vivo* by antisense oligonucleotides that target the splicing silencer in the same intron, a strategy not dissimilar to that used to correct the splicing defect in spinal muscular atrophy.[16,33] A key difference between the two antisense applications is the G-rich context in the vicinity of the *INS* antisense target.

Our results showed that the splicing silencer is surrounded by G-tracts prone to form RNA G4s *in vitro* and demonstrate the G4 formation in real time during *INS* transcription. Our data also revealed a wide range of G4 propensities for relatively short DNA and pre-mRNA segments *in vitro*. This folding landscape has great potential to generate a broad spectrum of dynamic pre-mRNA conformations that may help fine tune complex cellular processes such as co-transcriptional splicing. Although our results do not provide evidence for a role of G4 in this process *in vivo*, they strengthen the case for

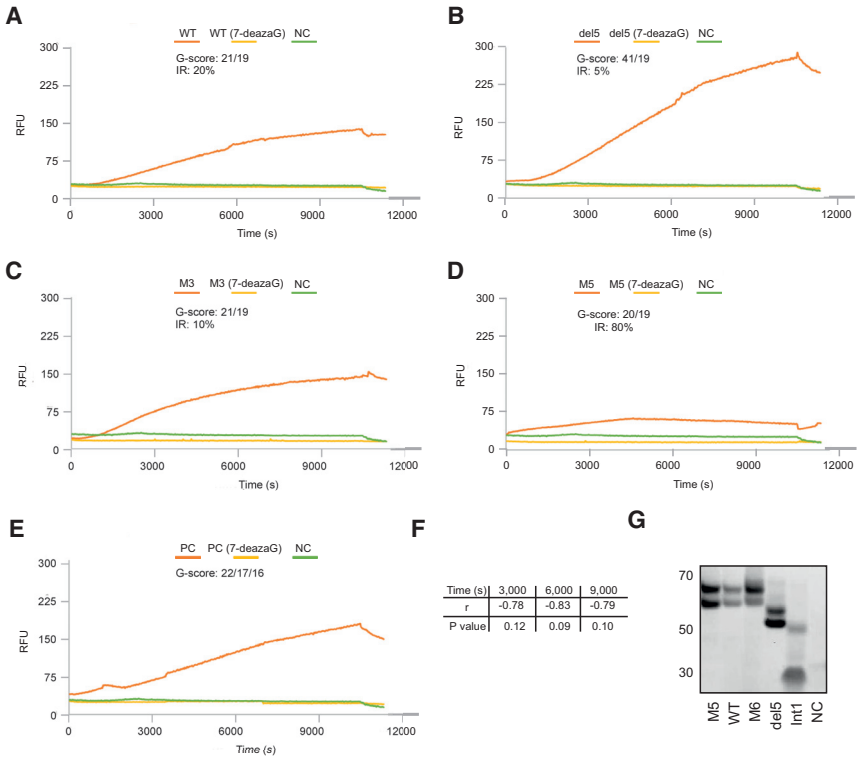

Figure 5. Real-Time ThT Monitoring of RNA G4 Formation during *INS* Transcription

(A–E) The fluorescence intensity of ThT was measured during transcription *in vitro* for 10,800 s and, after DNase addition, for an additional 900 s: (A), WT, (B) del5, (C) M3, (D) M5, and (E) positive control. Orange lines denote *INS*-derived sequences and a positive control (PC), yellow lines represent their 7-deazaG counterparts, and green lines denote no-template controls (NC). PC (Ampliscribe T7-Flash Kit) contained three putative G4s predicted by the QGRS Mapper.[28] (F) Pearson correlation coefficients between intron retention levels (%) and fluorescence intensity at the indicated time points. (G) Resolution of *INS*-derived pre-mRNAs (shown at the bottom) on a native polyacrylamide gel.

future pre-mRNA structural studies in insulin-expressing tissues, including pancreatic β cells.

Although previous studies point toward NMM selectivity for parallel G4 conformations, there have been reports showing fluorescence enhancements of NMM in the presence of anti-parallel G4s[30,35] and mixed-hybrid G4s.[36] However, it remains to be seen if the low level of induced NMM fluorescence by antiparallel structures could explain a low correspondence between ThT and NMM signal for *INS*-derived DNA oligonucleotides (Figures 1B and 1C).

Our results provide insights into the predictive value of G4 scores and highlight the importance of other factors in accurate G4 predictions than a mere presence or absence of G-tracts. Although G-runs are essential for G4 formation,[37–39] other factors are critical, including G-poor loop sequences,[40–43] which may contain accessible antisense targets. The ability of ThT to bind to pockets between adenine pairs[26] would explain the observed high ThT signal for adenine-rich controls, particularly for GA8 and GA12 (Figure 1B). The correlation with predicted G scores can be further reduced by non-homogeneous structures in solution, including hairpins in equilibrium with G4s (Figures 4 and 5).[16,30] Conversely, no or only a small fluorescence enhancement was found for oligonucleotides with predicted G4 structures, including Int2, Int3, and GT8/GA8 (Figure 1B). Finally, our data confirmed that G-rich oligoribonucleotides exhibit higher ThT fluorescence than their DNA counterparts and a stronger correlation with G4 prediction scores, consistent with a specificity of ThT for parallel conformations.[6,27]

MATERIALS AND METHODS

Synthetic Nucleic Acids

All oligonucleotides were purchased from Eurofins Genomics (Wolverhampton, UK). Their sequences are shown in Figures 1, 2, and 4; sequences of Q1 and double-stranded RNA oligonucleotides were 5′-CGG GGA AGG GCG CGG GCG CGG G-3′ (Q1) and 5′-AGU UCA AGG CGC CUU GAA CU-3′ (dsRNA), respectively. Oligonucleotides were purified by a high-purity, salt-free purification method and dissolved in RNase-free water (Gibco, Loughborough, UK). Concentrations of all oligonucleotides were determined spectrophotometrically using NanoDrop (Thermo Fisher Scientific, Waltham, MA). Their aliquots were stored at −20°C (DNA) or −80°C (RNA). Nucleic acids were visualized by staining the gel with 1× GelRed (Biotium, Cambridge, UK) for 30 min.

ThT Assay

2-(4-(Dimethylamino)phenyl)-3,6-dimethylbenzo[d]thiazol-3-ium chloride (ThT) dye was purchased from Sigma-Aldrich (Gillingham, UK) and used without further purification. The stock solution (5 mM) was prepared in DNase- and RNase-free water; aliquots were stored at −20°C. Fifty microliters of *INS*-intron-1-derived oligonucleotides and controls diluted to 20 μM were heated at 90°C for 5 minutes and cooled down to room temperature for 1 h at a rate of 1°C/min on a PeqSTAR 96 Universal Gradient thermocycler (PeqLab, Fareham, UK). ThT was added to a final concentration of 80 μM, which was derived by assay optimizations. ThT fluorescence intensities were measured at 508 nm (pH 7.2) using the FLx800 Microplate Fluorescence Reader (Biotek Instruments, Swindon, UK). The assay sensitivity was set to 90. *INS*-derived oligonucleotides and negative and positive controls were prepared and measured three times. The average 10 technical replicates were measured for each sample preparation. Positive controls included c-*myc* derived from the promoter sequence of the c-*myc* proto-oncogene, Plas24 derived from a plasmodium telomere and 21DNA from human *DBF4B* (de la Faverie et al.[21] and references therein).

NMM Assay

NMM (Frontier Scientific, Logan, UT, USA) was dissolved in DNase- and RNase-free water and kept at 4°C at a stock concentration of 100 μM. N-methyl mesoporphyrin IX (NMM) was added to oligonucleotides to a final concentration of 25 μM. Prior to NMM addition, G4 formation was promoted by heating the samples at 90°C for 5 minutes and cooling down to room temperature on a PeqSTAR 96 Universal Gradient thermocycler (PeqLab, Fareham, UK) for 1 h at a rate of 1°C/min. NMM fluorescence intensity was measured with a Varioskan multireader (Thermo Fisher Scientific, Waltham, MA, USA). Excitation wavelength was at 400 nm. Spectra were collected for a range between 435 and 740 nm.

ThT Monitoring during Transcription *In Vitro*

Transcription was carried out using PCR-generated templates amplified from the WT and mutated *INS* intron 1 minigenes described earlier.[14–16] PCR amplification was performed with primers 5′-att aat acg act cac tat aGG GCT CAG GGT TCC AGG-3′ and 5′-GCC CAG CCA CGT CCT CCC T-3′. The forward primer included a T7 promoter tag shown in the lower case. PCR was carried out using *Pfu* DNA polymerase (Promega, Southampton, UK) at a Mg^{2+} concentration of 2.4 mM at 95°C/5 min | 23 × (95°C/30 s | 61°C/45 s | 72°C/45 s, followed by an extension at 72°C for 2 min. To facilitate amplification of the GC-rich templates, DMSO was added to a final concentration of 5% (v/v). PCR products were visualized on 1.5% agarose gels, stained with GelRed (Biotium, Cambridge, UK) and gel purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA). RNA transcripts were generated, using the Ampliscribe T7-Flash Kit (Lucigen, Cambridge, UK) according to the manufacturer's instructions. As a negative control, we prepared transcription reactions with 7-deaza-GTP (Trilink Biotechnologies, San Diego, CA, USA), which inhibits Hoogsteen interactions and does not support G4 formation in nascent RNA transcripts.[32] After addition of ThT, reactions were initiated by adding T7 polymerase (Lucigen, Cambridge, UK). Fluorescence intensity was measured immediately after in a 96-well plate (Sigma-Aldrich, Gillingham, UK) by a Varioskan multireader (Thermo Fisher Scientific, Waltham, MA, USA) at 481 nm every 30 s. The excitation wavelength was 420 nm. After 3 h, RNase-free DNase I (New England Biolabs, Ipswich, MA, USA) was added to each well to stop the reaction, and fluorescence data were collected for another 15 min. RNA transcripts were recovered by phenol and chloroform extraction and ethanol precipitation and loaded onto a native 6% polyacrylamide gel.

Data Analysis

Statistical significance was determined with tests included in the GraphPad Prism (v. 7, GraphPad Software, La Jolla, CA, USA). The means and SDs were calculated from three biological replicates. The p values are denoted by asterisks in individual figure panels. For assessing normal distribution, we used the Shapiro-Wilk normality test. To determine statistical significance when comparing means of the fluorescence signals for tested oligonucleotides, we employed ANOVA, followed by Tukey's or Friedman's tests. For pairwise comparisons of data without normal distribution, we used the Wilcoxon-Mann-Whitney test.

AUTHOR CONTRIBUTIONS

A.L. and I.V. designed the study, A.L. performed the experiments, and A.L. and I.V. evaluated the data and prepared the manuscript. C.G.P. and J.W.H. contributed to data analysis and manuscript preparation. All authors approved the manuscript.

CONFLICTS OF INTEREST

The authors declare U.S. patent 9,714,422 B2 and a licensing agreement with ASOthera Pharmaceuticals, Inc.

ACKNOWLEDGMENTS

This study was supported by Diabetes UK (award 003962). We wish to thank Professor Ian Eperon (University of Leicester) for critical reading of the manuscript and Matt Brimmell and Melissa Doherty (University of Southampton) for technical help.

REFERENCES

1. Neidle, S., and Parkinson, G.N. (2003). The structure of telomeric DNA. Curr. Opin. Struct. Biol. *13*, 275–283.
2. Bardin, C., and Leroy, J.L. (2008). The formation pathway of tetramolecular G-quadruplexes. Nucleic Acids Res. *36*, 477–488.
3. Kwok, C.K., and Merrick, C.J. (2017). G-quadruplexes: prediction, characterization, and biological application. Trends Biotechnol. *35*, 997–1013.
4. Varizhuk, A., Ischenko, D., Tsvetkov, V., Novikov, R., Kulemin, N., Kaluzhny, D., Vlasenok, M., Naumov, V., Smirnov, I., and Pozmogova, G. (2017). The expanding repertoire of G4 DNA structures. Biochimie *135*, 54–62.
5. Yang, D., and Okamoto, K. (2010). Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med. Chem. *2*, 619–646.
6. Bugaut, A., and Balasubramanian, S. (2012). 5′-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. *40*, 4727–4741.
7. Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science *353*, aaf5371-1–aaf5371-8.
8. Huang, H., Zhang, J., Harvey, S.E., Hu, X., and Cheng, C. (2017). RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. Genes Dev. *31*, 2296–2309.
9. McCullough, A.J., and Berget, S.M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol. Cell. Biol. *17*, 4562–4571.
10. Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. USA *98*, 11193–11198.
11. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831–845.
12. Han, K., Yeo, G., An, P., Burge, C.B., and Grabowski, P.J. (2005). A combinatorial code for splicing silencing: UAGG and GGGG motifs. PLoS Biol. *3*, e158.
13. Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C.B. (2006). General and specific functions of exonic splicing silencers in splicing control. Mol. Cell *23*, 61–70.
14. Kralovicova, J., Gaunt, T.R., Rodriguez, S., Wood, P.J., Day, I.N., and Vorechovsky, I. (2006). Variants in the human insulin gene that affect pre-mRNA splicing: is -23HphI a functional single nucleotide polymorphism at IDDM2? Diabetes *55*, 260–264.
15. Kralovicova, J., and Vorechovsky, I. (2010). Allele-specific recognition of the 3′ splice site of *INS* intron 1. Hum. Genet. *128*, 383–400.
16. Kralovicova, J., Lages, A., Patel, A., Dhir, A., Buratti, E., Searle, M., and Vorechovsky, I. (2014). Optimal antisense target reducing INS intron 1 retention is adjacent to a parallel G quadruplex. Nucleic Acids Res. *42*, 8161–8173.

17. Type 1 Diabetes Genetics Consortium (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. *47*, 381–386.

18. Mohanty, J., Barooah, N., Dhamodharan, V., Harikrishna, S., Pradeepkumar, P.I., and Bhasikuttan, A.C. (2013). Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA. J. Am. Chem. Soc. *135*, 367–376.

19. Tong, L.L., Li, L., Chen, Z., Wang, Q., and Tang, B. (2013). Stable label-free fluorescent sensing of biothiols based on ThT direct inducing conformation-specific G-quadruplex. Biosens. Bioelectron. *49*, 420–425.

20. Gabelica, V., Maeda, R., Fujimoto, T., Yaku, H., Murashima, T., Sugimoto, N., and Miyoshi, D. (2013). Multiple and cooperative binding of fluorescence light-up probe thioflavin T with human telomere DNA G-quadruplex. Biochemistry *52*, 5620–5628.

21. Renaud de la Faverie, A., Guédin, A., Bedrat, A., Yatsunyk, L.A., and Mergny, J.L. (2014). Thioflavin T as a fluorescence light-up probe for G4 formation. Nucleic Acids Res. *42*, e65.

22. Bhasikuttan, A.C., and Mohanty, J. (2015). Targeting G-quadruplex structures with extrinsic fluorogenic dyes: promising fluorescence sensors. Chem. Commun. (Camb.) *51*, 7581–7597.

23. Du, Y.C., Zhu, L.N., and Kong, D.M. (2016). Label-free thioflavin T/G-quadruplex-based real-time strand displacement amplification for biosensing applications. Biosens. Bioelectron. *86*, 811–817.

24. Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., Wang, L., Liu, Y., Yu, L., Shi, Y., et al. (2016). Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes. Sci. Rep. *6*, 24793.

25. Wang, H., Peng, P., Liu, S., and Li, T. (2016). Thioflavin T behaves as an efficient fluorescent ligand for label-free ATP aptasensor. Anal. Bioanal. Chem. *408*, 7927–7934.

26. Liu, S., Peng, P., Wang, H., Shi, L., and Li, T. (2017). Thioflavin T binds dimeric parallel-stranded GA-containing non-G-quadruplex DNAs: a general approach to lighting up double-stranded scaffolds. Nucleic Acids Res. *45*, 12080–12089.

27. Joachimi, A., Benz, A., and Hartig, J.S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. Bioorg. Med. Chem. *17*, 6811–6815.

28. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, W676-82.

29. Bedrat, A., Lacroix, L., and Mergny, J.L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. *44*, 1746–1759.

30. Zhao, D., Dong, X., Jiang, N., Zhang, D., and Liu, C. (2014). Selective recognition of parallel and anti-parallel thrombin-binding aptamer G-quadruplexes by different fluorescent dyes. Nucleic Acids Res. *42*, 11612–11621.

31. Endoh, T., Rode, A.B., Takahashi, S., Kataoka, Y., Kuwahara, M., and Sugimoto, N. (2016). Real-time monitoring of G-quadruplex formation during transcription. Anal. Chem. *88*, 1984–1989.

32. Murchie, A.I., and Lilley, D.M. (1992). Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures. Nucleic Acids Res. *20*, 49–53.

33. Aartsma-Rus, A. (2017). FDA approval of nusinersen for spinal muscular atrophy makes 2016 the year of splice modulating oligonucleotides. Nucleic Acid Ther. *27*, 67–69.

34. Zhang, L., Nakayama, M., and Eisenbarth, G.S. (2008). Insulin as an autoantigen in NOD/human diabetes. Curr. Opin. Immunol. *20*, 111–118.

35. Kreig, A., Calvert, J., Sanoica, J., Cullum, E., Tipanna, R., and Myong, S. (2015). G-quadruplex formation in double strand DNA probed by NMM and CV fluorescence. Nucleic Acids Res. *43*, 7961–7970.

36. Sabharwal, N.C., Savikhin, V., Turek-Herman, J.R., Nicoludis, J.M., Szalai, V.A., and Yatsunyk, L.A. (2014). N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. FEBS J. *281*, 1726–1737.

37. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr., and Blackburn, E.H. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. Cell *51*, 899–908.

38. Marathias, V.M., and Bolton, P.H. (1999). Determinants of DNA quadruplex structural type: sequence and potassium binding. Biochemistry *38*, 4355–4364.

39. Granqvist, L., and Virta, P. (2016). Characterization of G-quadruplex/hairpin transitions of RNAs by ^{19}F NMR spectroscopy. Chemistry *22*, 15360–15372.

40. Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. J. Am. Chem. Soc. *126*, 16405–16415.

41. Rachwal, P.A., Brown, T., and Fox, K.R. (2007). Sequence effects of single base loops in intramolecular quadruplex DNA. FEBS Lett. *581*, 1657–1660.

42. Bugaut, A., and Balasubramanian, S. (2008). A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. Biochemistry *47*, 689–697.

43. Zhang, A.Y., Bugaut, A., and Balasubramanian, S. (2011). A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. Biochemistry *50*, 7251–7258.