G-Quadruplexes

International Edition: DOI: 10.1002/anie.201603562
German Edition: DOI: 10.1002/ange.201603562

Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA

*Chun Kit Kwok, Aleksandr B. Sahakyan, and Shankar Balasubramanian**

Abstract: *RNA G-quadruplex (rG4) structures are of fundamental importance to biology. A novel approach is introduced to detect and structurally map rG4s at single-nucleotide resolution in RNAs. The approach, denoted SHALiPE, couples selective 2′-hydroxyl acylation with lithium ion-based primer extension, and identifies characteristic structural fingerprints for rG4 mapping. We apply SHALiPE to interrogate the human precursor microRNA 149, and reveal the formation of an rG4 structure in this non-coding RNA. Additional analyses support the SHALiPE results and uncover that this rG4 has a parallel topology, is thermally stable, and is conserved in mammals. An in vitro Dicer assay shows that this rG4 inhibits Dicer processing, supporting the potential role of rG4 structures in microRNA maturation and post-transcriptional regulation of mRNAs.*

RNA G-quadruplexes (rG4s) appear to be important in gene regulation and disease.[1] This structural motif comprises G-quartets (Figure 1 A) with connecting loops, and can be stabilized by cations, especially potassium (K^+). Recently, rG4s have been visualized in human cells,[2] and ligands, such as pyridostatin (PDS)[3] (Figure 1 B), have been shown to further stabilize rG4 structures.[2a]

Our understanding of rG4 structure and function is still in its infancy. Biophysical characterization of rG4s has typically involved circular dichroism (CD) spectroscopy,[4] UV-thermal melting,[5] and NMR spectroscopy,[6] which report rG4 features, but largely without considering the effects of flanking sequences and an extended sequence context on rG4 formation. In-line probing can address these shortcomings and identifies rG4s;[7] however, the long experimental time (two-day reaction) poses some practical limitations. Reverse transcriptase stalling (RTS)[8] can map the 3′-end of the rG4 in a short time (15-minute reaction), but does not provide nucleotide resolution details of the G-quartets and loops.

SHAPE[9] (selective 2′-hydroxyl acylation analyzed by primer extension) exploits differential, structure-dependent kinetics of acylation of 2′-OH of RNA, which is measured by reverse transcriptase stop at the 2′-O adduct. SHAPE has mapped various RNA structures,[10] including motifs such as RNA hairpins, but not rG4s. Herein, we show a new method to map rG4s that overcomes limitations of the classical SHAPE method. Specifically, we introduce lithium ion (Li^+)-based primer extension (LiPE) with 2-methylnicotinic acid imidazolide (NAI) to map rG4s at single nucleotide resolution (Figure 1 C). Importantly, we apply it to reveal rG4 formation in biologically important RNAs, and report the role of rG4s in Dicer processing.

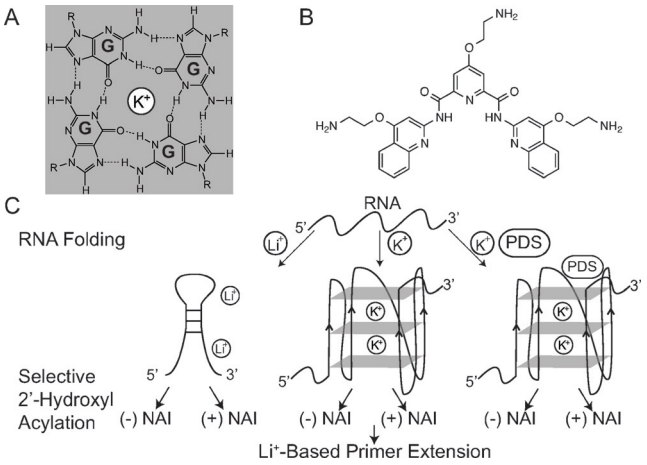

Figure 1. Structure probing of rG4s using SHALiPE. Chemical structure of A) a G-quartet and B) pyridostatin (PDS). C) Overview of SHALiPE. RNA is folded under Li^+, K^+, or K^+ + PDS conditions, followed by selective 2′-hydroxyl acylation (SHA) using NAI. Li^+-based primer extension (LiPE) is conducted to analyze the NAI-modified RNA. Controls are performed with no NAI.

We first evaluated the current SHAPE approach on an rG4 structure, using a construct comprising 5′ and 3′ stem loops (SL)[11] and the telomeric *TERRA*[12] rG4, UUA**GGG**UUA**GGG**UUA**GGG**UUA**GGG**UUA (Supporting Information, Table S1 and Figure S1). Using NAI-acylation (5-minute reaction),[13] we observed that patterns of NAI-induced modifications are obscured by strong RTS by *TERRA* rG4 formation, in the standard K^+-containing PE buffer (Figure 2, lanes 1–6, green). This suggests that K^+ should be avoided in the PE buffer[14] to enable SHAPE analysis of rG4s.

As rG4 structures are less stable in Li^+,[15] we prepared and evaluated a Li^+-based PE buffer to reduce rG4 stalling and unmask patterns of rG4-dependent 2′-OH acylation. When the reaction of RNA with NAI was performed in the presence of Li^+, the 2′-OH of most nucleotides in the *TERRA* rG4 region were strongly modified (Figure 2, lane 10), indicating that the rG4 region is unstructured. In contrast,

[*] Dr. C. K. Kwok, Dr. A. B. Sahakyan, Prof. S. Balasubramanian
The University of Cambridge, Department of Chemistry
Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: sb10031@cam.ac.uk

Supporting information for this article can be found under: http://dx.doi.org/10.1002/anie.201603562.

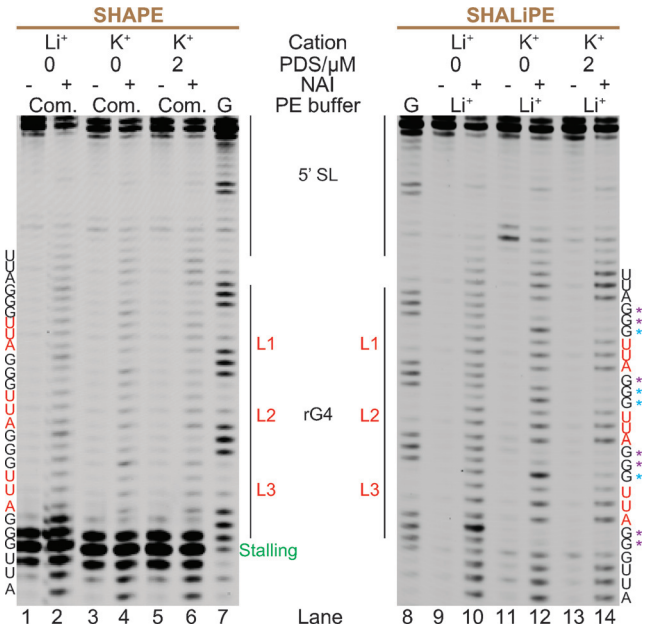

Figure 2. SHALiPE development and analysis. *TERRA* RNA was probed with NAI under Li^+, K^+, and K^+ + PDS conditions, followed by primer extension (PE) using either the SSIII commercial (com.) K^+-containing PE buffer[14] (left gel) or home-made Li^+-containing PE buffer (right gel). On the left, intense stalling was observed (lanes 1–6), caused by formation of *TERRA* rG4 (stabilized by K^+ in the commercial PE buffer) in the PE step. On the right, stalling was reduced using the Li^+-based PE buffer. Lanes 9–14 show (−) and (+) NAI signals under Li^+, K^+, and K^+ + PDS conditions. *TERRA* rG4 was unstructured in Li^+ conditions. Under physiological K^+ conditions, *TERRA* rG4 displayed distinct NAI profiles (low NAI signals in several Gs involved in rG4; purple asterisks, see lanes 10 and 12). Blue asterisks indicate Gs that were protected upon PDS addition (lanes 12 and 14). The loops (L1, L2, and L3) and loop nucleotides (UUA) are shown in red. Lanes 7–8 show sequencing of G. 5′ SL, 5′ stem loop (Supporting Information, Figure S1).

NAI reaction in the presence of K^+, led to protection of some Gs in the rG4 region (Figure 2, lanes 10 and 12, purple asterisks), suggestive of rG4 formation. Interestingly, the modifications at the remaining Gs in the rG4 motif were more intense in K^+ compared to Li^+ (Figure 2, lanes 10 and 12, blue asterisks), but more protected in the presence of both K^+ and the rG4 stabilizing ligand, namely PDS (K^+ + PDS) (Figure 2, lanes 12 and 14, blue asterisks). These Gs were mostly located at the 3′-end of each G-tract. Similar findings were observed for *TERC* rG4 and *NRAS* rG4 (Supporting Information, Figures S2, S3), suggesting that the 2′-OHs of the 3′-G-quartet are more flexible than those of the 5′-G-quartet. These data show that SHALiPE provides a distinctive pattern for rG4 structure that can be used to fine map rG4s, along with other RNA structures (Supporting Information, Figure S4) at single nucleotide resolution.

RNA structures are dynamic and hairpin to G-quadruplex conformational transition[16] may have regulatory roles in biological processes. We surveyed the miRBase[17] to identify precursor microRNA (pre-miRNA) hairpins that contain putative rG4s (Supporting Information, Table S2), and selected a conserved candidate (Supporting Information, Table S3), *pre-miRNA 149*, for further investigation. The processing of *pre-miRNA 149* by Dicer produces *miR149* and *miR149**, both of which have been shown to function as oncogenic regulators in cancers.[18]

Using comparative sequence analysis, we identified a conserved rG4 sequence that partially overlaps with *miRNA 149** (Figure 3A). We confirmed rG4 formation in the putative rG4 sequence by CD and UV-melting spectroscopic analyses (Figure 3B; Supporting Information, Figure S5). Under 150 mM K^+ conditions, the CD spectrum of the wildtype rG4 sequence, but not the mutated sequence, showed a distinct CD profile consistent with a parallel rG4 topology[4] (Figure 3B; Supporting Information, Table S1). UV-melting analysis of the wildtype rG4 sequence displayed a hypochromic

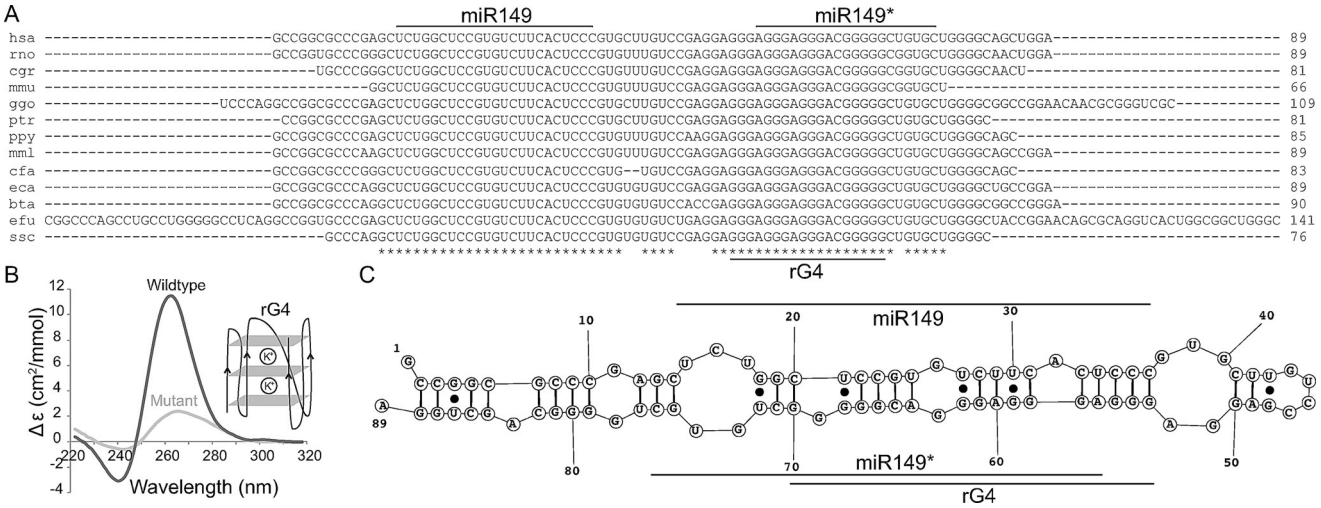

Figure 3. Computational analysis and biophysical characterization of a putative rG4 in human *pre-miRNA 149*. A) Comparative sequence analysis of *pre-miRNA 149*. The *miR149*, *miR149**, and putative rG4 are highlighted. Conserved nucleotides are marked by asterisks and the length of RNAs are shown. The name of the species, presented here with three-letter codes, can be found in the Supporting Information, Table S3. B) CD spectrum of wildtype rG4 sequence under K^+ condition shows a characteristic signature that indicates formation of a parallel topology rG4 structure. Mutant shows no such signature. The *Y* axis is molar ellipticity per nucleotide ($\Delta\varepsilon$). C) The secondary structure model for the human *pre-miRNA 149* is derived using the sequences in (A) and TurboFold.[19]

shift at 295 nm (Supporting Information, Figure S5), a hallmark for rG4 formation.[5] Furthermore, the melting temperature for the rG4 sequence under the 150 mM K^+ condition was found to be > 90 °C, suggestive of a highly thermostable rG4 under physiological K^+ conditions.

Next, we predicted the secondary structure of the human *pre-miRNA 149* by TurboFold,[19] which takes into consideration the sequences of orthologues and generates a consensus RNA structural model (Figure 3C). Since the rG4 partially overlaps with the canonical hairpin structure in *pre-miRNA 149* (Figure 3C), we hypothesize that the formation of an rG4 structure may affect the folding of the hairpin structure in this pre-miRNA.

We applied SHALiPE to probe the structure of *pre-miRNA 149* under Li^+, K^+, and K^+ + PDS conditions (Figure 4A). The results revealed a number of important findings. First, NAI modification (reactivity) profiles for Li^+ versus K^+ and Li^+ versus K^+ + PDS were weakly correlated, with Pearson's correlation coefficients (PCC) of 0.10 and 0.19 respectively (Figure 4B, left and middle plots), suggesting that the major RNA conformations under Li^+ and K^+ (or K^+ + PDS) conditions are different.

Second, we found that in Li^+ (Figure 4A, lanes 1–2), NAI modifications mapped consistently to the predicted hairpin structure (Figure 4A,C, green asterisks). In contrast, under K^+ and K^+ + PDS conditions (Figure 4A, lanes 3–6), NAI modifications did not agree with the hairpin structure. For example, A56, A60, and A64 were strongly modified by NAI (Figure 4A, lanes 3–4), despite being assigned as base-paired nucleotides in the hairpin structure (Figure 4C). To verify the SHALiPE results, a nucleobase-specific reagent dimethyl sulfate (DMS) was used, which modifies the Watson–Crick face of As (N_1A) and Cs (N_3C)[20] in a manner that stalls reverse transcriptase. As such, DMSLiPE provides a measure of nucleobase-specific DMS reactivity with RNA. Similar to NAI, we observed that A56, A60, and A64 were strongly modified by DMS under K^+ and K^+ + PDS conditions but not in Li^+ conditions (Supporting Information, Figure S6), suggesting that A56, A60, and A64 are not base-paired and are the loop nucleotides of the rG4 (Figure 4A,D). Our results support that the major RNA conformation under K^+ and K^+ + PDS conditions contains an rG4 (Figure 4D).

Furthermore, we performed SHALiPE on a mutant construct that cannot form an rG4, and found that the NAI modification profiles were almost identical under Li^+, K^+, and K^+ + PDS conditions (Supporting Information, Figure S7). Also, the NAI modifications mapped well to the predicted structure of this construct (Supporting Information, Figure S7). This verified that the change in NAI profiles observed in the natural sequence (Figure 4A) was indeed due to rG4 formation.

Third, we observed that the PCC for NAI reactivity values under K^+ and K^+ + PDS conditions was 0.65 (Figure 4B, right plot). Visually, three nucleotides (G63, G55, and G59) stood out, and by removing them from the correlation analysis yielded a PCC of 0.95 (Figure 4B, right plot), similar to the

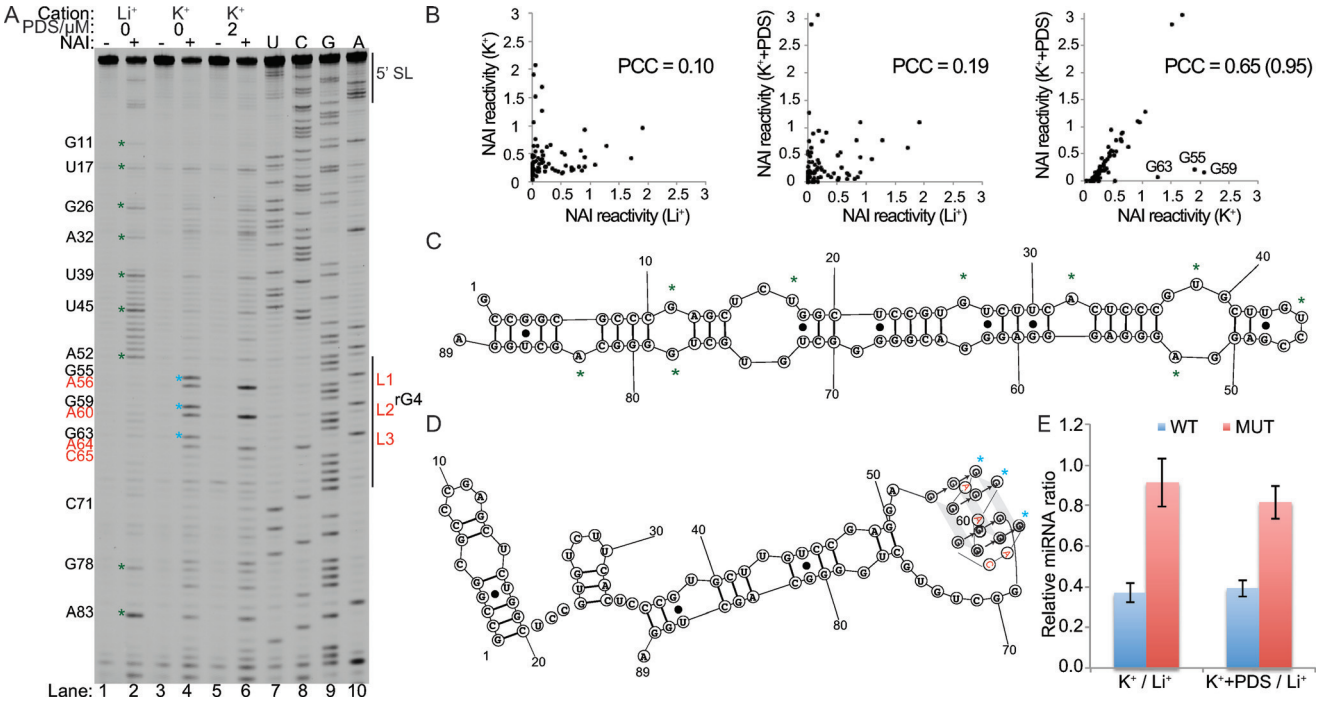

Figure 4. SHALiPE and Dicer cleavage assays on human *pre-miRNA 149* reveal an rG4 that inhibits in vitro Dicer processing. A) SHALiPE on *pre-miRNA 149* under Li^+, K^+, and K^+ + PDS conditions (lanes 1–6). Lanes 7–10 show sequencing of U, C, G, and A respectively. The rG4 loops are in red. G that are protected with the addition of PDS (G55, G59, G63) are marked with blue asterisks. B) The PCC of the normalized NAI reactivity for K^+ versus Li^+ (left), K^+ + PDS versus Li^+ (middle), and K^+ + PDS versus K^+ (right). The PCC increases from 0.65 to 0.95 by removing G55, G59 and G63 (right). C) The hairpin conformation of *pre-miRNA 149*. Some nucleotides are marked with green asterisks. D) The rG4-containing conformation of *pre-miRNA 149*. G55, G59, and G63 are marked with blue asterisks. E) Results of in vitro Dicer assay on *pre-miRNA 149* wildtype (WT) and mutant (MUT). Relative miRNA ratio (K^+/Li^+ or K^+ + PDS/Li^+) are reported. Values are from 3 biological replicates and error bars depict standard deviations. The constructs for the Dicer assay do not contain the 5′ and 3′SL (Supporting Information, Table S1).

PCC of 0.93 for DMS reactivity (only considers As and Cs)[20] (Supporting Information, Figure S6). Interestingly, these three nucleotides were the 3′-G of each G-tract in the rG4 (Figure 4D, blue asterisks). They were highly modified by NAI under K^+, and protected with the addition of PDS (Figure 4A, lanes 4 and 6), as for the other rG4 examples we tested (Figure 2; Supporting Information, Figures S2, S3). Different acylation reagents and rG4 ligands yielded the same results (Supporting Information, Figure S8), suggesting that this feature is not reagent- or rG4 ligand-dependent, but rather due to the flexible nature of the 3′-G-quartet and/or the unique conformation of the 3′-Gs of the rG4, making them susceptible for 2′-OH acylation attack.

Taken together, these results provide strong evidence for the formation of an rG4 under K^+ and K^+ + PDS conditions (Figure 4D), whereas a hairpin conformation predominates under Li^+ conditions (Figure 4C) in the *pre-miRNA 149* wildtype. It is possible that two (or more) RNA conformations exist under K^+ and K^+ + PDS conditions.

To assess the role of the rG4 structure in this *pre-miRNA 149*, we performed an in vitro Dicer cleavage assay on the wildtype and mutant constructs (without 5′ and 3′ SL) under Li^+, K^+, and K^+ + PDS conditions (Figure 4E; Supporting Information, Figure S9, Table S1). Our results showed that the rG4 inhibits Dicer processing activity. The yield of miRNA under K^+ and K^+ + PDS conditions were 2.5–2.7-fold less than under the Li^+ condition for the wildtype (Figure 4E, blue bars, 0.37 ± 0.05 for K^+/Li^+ and 0.39 ± 0.04 for K^+ + PDS/Li^+), in contrast to 1.1–1.2-fold for the mutant (Figure 4E, red bars, 0.91 ± 0.12 for K^+/Li^+ and 0.82 ± 0.08 04 for K^+ + PDS/Li^+). The repressive role of the rG4 in *pre-miRNA 149* is consistent with recently reported cases of *pre-miRNA 92b* and *pre-miRNA let-7e*,[21] suggesting a broader role for rG4s in miRNA maturation. Our bioinformatics analysis shows 85 putative rG4s in human pre-miRNAs (using $G_{3+}L_{1\text{-}12}$-based motif, see the Supporting Information) and many others across organisms (Supporting Information, Table S2), suggested the phenomenon may be widespread.

In summary, SHALiPE (and DMSLiPE) has enabled fine mapping of rG4 structures in RNAs. The approach has numerous benefits over prior methods and we have used it to reveal an rG4 structure in the biologically important *pre-miRNA 149*, and showed that the rG4 inhibits Dicer processing in vitro. Our findings provide insights into the biological roles of rG4 structures and the approach may ultimately have utility in transcriptome-wide probing of rG4s.

Acknowledgements

This work is supported by a European Research Council Advanced grant (No. 339778). C.K.K. received some support from the Croucher Foundation. A. Sahakyan was supported by a Herchel Smith Fellowship. We thank V. Chambers, Dr. B. Herdy, Dr. L. K. M. Chan, and Dr. M. Di Antonio for discussions.

Keywords: Dicer processing · G-quadruplexes · precursor miRNA · RNA structure · structure probing

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 8958–8961
Angew. Chem. **2016**, *128*, 9104–9107

[1] a) S. Millevoi, H. Moine, S. Vagner, *WIREs RNA* **2012**, *3*, 495–507; b) R. Simone, P. Fratta, S. Neidle, G. N. Parkinson, A. M. Isaacs, *FEBS Lett.* **2015**, *589*, 1653–1668.
[2] a) G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 75–80; b) A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J. M. Perrier-Cornet, J. M. Wong, D. Monchaud, *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525.
[3] R. Rodriguez, S. Muller, J. A. Yeoman, C. Trentesaux, J. F. Riou, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759.
[4] M. Vorlíčková, I. Kejnovská, J. Sagi, D. Renčiuk, K. Bednářová, J. Motlová, J. Kypr, *Methods* **2012**, *57*, 64–75.
[5] J.-L. Mergny, A.-T. Phan, L. Lacroix, *FEBS Lett.* **1998**, *435*, 74–78.
[6] M. Webba da Silva, *Methods* **2007**, *43*, 264–277.
[7] a) J. D. Beaudoin, R. Jodoin, J. P. Perreault, *Methods* **2013**, *64*, 79–87; b) C. K. Kwok, Y. Ding, S. Shahid, S. M. Assmann, P. C. Bevilacqua, *Biochem. J.* **2015**, *467*, 91–102.
[8] C. K. Kwok, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2015**, *54*, 6751–6754; *Angew. Chem.* **2015**, *127*, 6855–6858.
[9] K. A. Wilkinson, E. J. Merino, K. M. Weeks, *Nat. Protoc.* **2006**, *1*, 1610–1616.
[10] K. M. Weeks, D. M. Mauger, *Acc. Chem. Res.* **2011**, *44*, 1280–1291.
[11] This design allows information near the 5′- and 3′-end of the sequence of interest to be retrieved. For gel-based analysis, the 5′-end of the RNA has band compression and intense full length band, whereas the 3′-end of the RNA requires a primer bindng site.
[12] B. Luke, J. Lingner, *EMBO J.* **2009**, *28*, 2503–2510.
[13] R. C. Spitale, P. Crisalli, R. A. Flynn, E. A. Torre, E. T. Kool, H. Y. Chang, *Nat. Chem. Biol.* **2013**, *9*, 18–20.
[14] Reverse transcriptases, for example, Superscript, AMV, Protoscript, Thermoscript, Maxima, and Tth Pol, contain 150–375 mM K^+ in their 5X primer extension buffers.
[15] S. Neidle, S. Balasubramanian, *Quadruplex nucleic acids*, *Vol. 7*, Royal Society of Chemistry, Cambridge, **2006**.
[16] A. Bugaut, P. Murat, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19953–19956.
[17] A. Kozomara, S. Griffiths-Jones, *Nucleic Acids Res.* **2014**, *42*, D68–73.
[18] a) L. Jin, W. L. Hu, C. C. Jiang, J. X. Wang, C. C. Han, P. Chu, L. J. Zhang, R. F. Thorne, J. Wilmott, R. A. Scolyer, P. Hersey, X. D. Zhang, M. Wu, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15840–15845; b) A. Bischoff, B. Huck, B. Keller, M. Strotbek, S. Schmid, M. Boerries, H. Busch, D. Muller, M. A. Olayioye, *Cancer Res.* **2014**, *74*, 5256–5265; c) S. H. Chan, W. C. Huang, J. W. Chang, K. J. Chang, W. H. Kuo, M. Y. Wang, K. Y. Lin, Y. H. Uen, M. F. Hou, C. M. Lin, T. H. Jang, C. W. Tu, Y. R. Lee, Y. H. Lee, M. T. Tien, L. H. Wang, *Oncogene* **2014**, *33*, 4496–4507.
[19] A. Harmanci, G. Sharma, D. Mathews, *BMC Bioinf.* **2011**, *12*, 108.
[20] DMS also modifies the unprotected, Hoogsteen face of Gs (N7G); however, they are undetected in reverse transcription.
[21] a) G. Mirihana Arachchilage, A. C. Dassanayake, S. Basu, *Chem. Biol.* **2015**, *22*, 262–272; b) S. Pandey, P. Agarwala, G. G. Jayaraj, R. Gargallo, S. Maiti, *Biochemistry* **2015**, *54*, 7067–7078.

Received: April 12, 2016
Published online: June 29, 2016

