Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator

Rhianna K. Morgan¹, Harshul Batra¹, Vanessa C. Gaerig², Jennifer Hockings¹, and Tracy A. Brooks^{1*}

¹School of Pharmacy, Department of BioMolecular Sciences, Division of Pharmacology, University of Mississippi, University, MS, 38677

²College of Pharmacy, Department of Pharmacology and Toxicology, The University of Arizona, Tucson, AZ 85721, USA

*Corresponding Author

Mail address: 307 Faser Hall, University, Mississippi 38677

Telephone: 662-915-2309

Fax: 662-915-5148

Email: tabrooks@olemiss.edu

Running Title: New silencing G-quadruplex identified in the kRAS promoter

Keywords: G-quadruplex, kRAS, transcriptional regulation, cancer therapeutics

Financial Support: Department of Defense PRCRP grant (to Brooks) CA130229, Startup funds from the University of Mississippi (to Brooks)

Abstract

kRAS is one of the most prevalent oncogenic aberrations. It is either upregulated or mutationally activated in a multitude of cancers, including pancreatic, lung, and colon cancers. While a significant effort has been made to develop drugs that target kRAS, their clinical activity has been disappointing due to a variety of mechanistic hurdles. The presented works describe a novel mechanism and molecular target to downregulate kRAS expression – a previously undescribed G-quadruplex (G4) secondary structure within the proximal promoter acting as a transcriptional silencer. There are three distinct guanine-rich regions within the core kRAS promoter, including a previously examined region ($G4_{\text{near}}$). Of these regions, the most distal region does not form an inducible and stable structure, whereas the two more proximal regions (termed near and mid) do form strong G-quadruplexes. $G4_{\text{near}}$ is predominantly a tri-stacked structure with a discontinuous guanine run incorporated; $G4_{\text{mid}}$ consists of seven distinct runs of continuous guanines and forms numerous competing isoforms, including a stable three-tetrad stacked mixed parallel and antiparallel loop structures with longer loops of up to 10 nucleotides. Comprehensive analysis of the regulation of transcription by higher order structures has revealed that the guanine-rich region in the middle of the core promoter, termed $G4_{\text{mid}}$, is a stronger repressor of promoter activity than $G4_{\text{near}}$. Using the extensive guanine-rich region of the kRAS core promoter, and particularly the $G4_{\text{mid}}$ structure, as the primary target, future drug discovery programs will have potential to develop a potent, specifically targeted small molecule to be used in the treatment of pancreatic, ovarian, lung, and colon cancers.

Keywords: G-quadruplex, kRAS, transcriptional control, cancer therapeutics

1. Introduction

The *kRAS* gene, located in chromosome 12 at p12.1 [1], encodes for the p21^{RAS} (kRAS) protein that participates in the Raf-MAP kinase pathway powering cell growth and apoptosis. Activated kRAS predominantly signals through the mitogen-activated protein kinase pathway (Raf/MEK/ERK), Phosphoinositide 3'-kinases, RalGEF, Phospholipase C, and MEKK1 [2]. The best characterized activation pathway for *kRAS* is via tyrosine kinase receptors like EGFR. Mutations in RAS proteins are found in approximately 30% of all human tumors, with *kRAS* being the most frequently mutated isoform [2, 3]; such mutations render the protein constitutively active.

Single point mutations of the *kRAS* gene, frequently occurring in codons 12, 13, and 61, abolish GAP-induced GTP hydrolysis through steric hindrance (G12 and G13) or by interfering with coordination of a water molecule necessary for GTP hydrolysis (Q61)[4]. The highest incidence of mutational activation occurs in colorectal, ovarian, and lung cancers, and most notably in >95% of pancreatic adenocarcinomas [1, 3]. kRAS mutations have been associated with increased tumorigenicity and poor prognosis[2]. Endogenous expression of mutant kRAS^{G12D} was capable of inducing pancreatic intraepithelial neoplasias (PIN) in a mouse model [5]. In the absence of a mutation, increased kRAS activity in human tumors has been shown to be the result of gene amplification, overexpression, or increased upstream activation[2]. Targeting kRAS expression or activity is well validated to inhibit tumor cell proliferation.

While there have been many attempts to mitigate mutant kRAS activity, including farnesyltransferase inhibitors, Raf kinase inhibitors, and MEK inhibitors, no clinically relevant agent that specifically targets kRAS currently exists [1]. Studies have shown that a reduction in kRAS expression in cancerous cells, by antisense, miRNA or siRNA oligonucleotides, halts proliferation and leads to cellular death [6, 7]. Furthermore, the inhibition of activated kRAS has been shown to revert malignant cells to a non-malignant phenotype, and cause tumor regression both *in vitro* and *in vivo*[2, 8]. Transcriptional or translational downregulation of kRAS has been validated as a novel therapeutic approach [1, 6, 7], with great potential to succeed where previous efforts focused on modulating *kRAS* signaling did not [3]. However, achieving this downregulation therapeutically is currently hampered by the lack of a molecular target. The works presented highlight a viable new target to fill this gap.

The core promoter region of *kRAS* is encompassed within the region from +50 bp through -510 bp, in relation to the transcriptional start site (TSS). The DNA within this region is highly G/C-rich (~75%), putatively capable of forming higher order non-B-DNA structures, and contains two nuclease hypersensitivity elements [9-11]. Such G/C-rich regions preferentially cluster around the transcriptional start site throughout the genome [12], with a high prevalence in oncogenic promoters [13]. Negative superhelicity induced by transcription can promote local unwinding of these G/C-rich regions of DNA, which allows for the formation of secondary structures known as G-quadruplexes (G4s).

G4s form from guanine-rich sequences by binding together four guanines in a planar fashion by Hoogsteen hydrogen bonds. Three or more planar guanine tetrads are vertically stacked either with multiple strands interacting in an intermolecular formation *ex vivo*, or with one strand folding upon itself in the biologically relevant intramolecular isoform. Loops connecting the runs of three or more continuous guanines can be in either the parallel or antiparallel configuration, and are typically 1 – 9 base pairs in length, although they have been described up to 26 base pairs. G4s have received much attention recently in the cancer community as their prevalence within the genome is notably higher in oncogenic promoters. Formation of G4s in DNA has been shown to modulate transcription, and in RNA modulates translation. Formation of G4s in DNA has been recently shown to clearly form *in vivo*, where it modulates transcription [14-19]. Their more unique, non-B-DNA, structure and potential ability to regulate the transcription of a host of oncogenes make G-quadruplexes an attractive drug target.

The region of the *kRAS* promoter from -129 to -160 has been previously examined [20-23], and various G-quadruplex formations were shown to exist in equilibrium. Examination of the *kRAS* promoter revealed an extensive region of G-rich DNA covering almost 300 bases from the TSS, and including a total of three putative G4-forming regions separated by 17 and 12 nucleotides in the 5'–3'-direction, respectively. These include the aforementioned previously described near- (five G-tracts over 30 bp), and the newly described mid- (seven G-tracts over 53 bp), and far- (4 G-tracts over 35 bp) regions. Each of these regions' G-quadruplex formations and function within the *kRAS* promoter are explored through biophysical studies, and *in vitro* with endogenous regulation in Panc-1 pancreatic cancer cells, as well as with a series of luciferase plasmid constructs. Cumulatively from these studies, it is clear that a major silencing

G-quadruplex is formed from the mid-G-quadruplex forming region of the *kRAS* promoter, and that it is this structure that represents a promising new molecular target. These works present a previously undescribed regulatory region of the core *kRAS* promoter, and initial step in the drug discovery process for new compounds aimed at targeting kRAS. The development of compounds that are more specific and potent stabilizers of the unique mid-G-quadruplex have great potential as anti-cancer therapeutics to advance the care of pancreatic, lung, and colon cancer patients.

2. Materials and Methods

2.1. Materials: All oligonucleotides (Table 1) were purchased from Operon (Huntsville, AL). Acrylamide/bisacrylamide (29:1) solution and ammonium persulfate were purchased from Bio-Rad laboratories (Hercules, CA), and *N,N,N',N'*-tetramethylethylenediamine was purchased through Fisher Scientific (Pittsburgh, PA). *Taq* DNA polymerase, T4 polynucleotide kinase, pRL-SV40 and pGL4.17 plasmids, dual luciferase assay kits were purchased from Promega (Madison, WI). ^{32}P -ATP was purchased from NEN Dupont. All other chemicals, unless otherwise noted, were purchased from Sigma-Aldrich (St. Louis, MO).

2.2. Circular Dichroism (CD): CD spectra and thermal stability of all sequences (5 μM in 50 mM Tris-HCl, pH 7.4) in the absence and presence of KCl (up to 100 mM) and Acetonitrile (ACN, up to 40% [24-29]), as indicated in the text, were recorded on an Olis DSM-20 Spectrophotometer fitted with a CD 250 Peltier cell holder (Bogart, GA) from 225-350 nm with scanning time as a function of high volts (<1 s/nm). Thermal stability was determined through collection of spectra (225-350 nm) from 20-100 $^{\circ}\text{C}$ (spectra recorded each 7 $^{\circ}\text{C}$, sample held at temp for 1 min before spectral scan) and single value decomposition (SVD) analysis was performed [30], from which non-linear regression was used to determine the T_{M} . All spectra were baselined for signal contributions from the buffer.

2.3. Electrophoretic mobility shift assay (EMSA): FAM-labeled oligonucleotides were denatured by heating to 95 $^{\circ}\text{C}$ for 10 min and then slowly cooled in 50 mM Tris-HCl, pH 7.4, with or without 100 mM KCl alone or with 40% ACN to induce G4 formation. After the addition of non-denaturing loading dye, samples were loaded on a 10% native polyacrylamide (29:1

acrylamide:bisacrylamide) gel, which was 50V; the gel was visualized under blue light LED using a FotoDyne Investigator FX Imager.

2.4. Oligonucleotide end-labeling and purification: As indicated, DNA oligomers were 5'-end-labeled with [γ - ^{32}P]ATP with T4 polynucleotide kinase for 1 h at 37 °C. The reaction was stopped by heating the samples to 90 °C for 5 min; the 5'-end-labeled DNA was purified with a Bio-Spin 6 micro-chromatography column (Bio-Rad Laboratories).

2.5. Dimethyl Sulfate (DMS) footprinting: 5'-end P- or 5'-FAM-labeled DNA oligonucleotides were denatured by heating to 95 °C for 5 min and then slowly cooled at 4 °C in 50 mM Tris-HCl buffer with or without monovalent cations. Following the addition of 1 μg of calf thymus DNA, the solutions were subjected to dimethyl sulfate (0.3% DMS in 1% ethanol) for up to 18 minutes. Each reaction was quenched 0.3 M NaOAc and 0.2 M β -mercaptoethanol. The reactions were subjected to a preparative gel and each band of interest was excised and eluted. After ethanol precipitation and treatment with 1 M piperidine, the DNA was dried and washed with water, the resuspended in a 95% formamide loading dye, and heated to 95 °C for 5 min before snap-cooling on ice. The cleaved products were separated on a 16% sequencing gel. Gels of G4_{near} with ^{32}P -DNA were then dried and exposed to a phosphorscreen and imaging was performed with a Storm 820 phosphorimager. Footprinting gels for FAM-labeled G4_{mid} were imaged under blue LED excitation and filtered through a blue emission filter on a FotoDyne Investigator FX Imager (Hartland, WI). Due to the size of the gels, they were imaged by the top and bottom regions, which were aligned using Adobe Illustrator; the images differentially obtained are distinguished by boxed outline in Figure 3.

2.6. *FRET Melt Assay*: Thermodynamic stabilities of G4DNA were performed using a fluorescence resonance energy transfer(FRET) melting assay[31]. Oligonucleotides were dual-labeled with 6-FAM on the 5' end and a TAMRA quencher on the 3' end (Operon for G4_{near} and Midland Certified Reagent Company, Midland, TX for G4_{mid}). Stock concentrations of DNA were made in autoclaved, nuclease-free water at an approximate concentration of 100 μ M. Assay concentrations and conditions were as reported; briefly, 200 nM of dual-labeled DNA was diluted in a 10 mM lithium cacodylate buffer (pH 7.2) supplemented with 10 mM potassium chloride, heated to 95 °C for 10 min, and allowed to anneal to room temperature slowly. Annealed DNA probe was then plated with 1 μ M compound in a 96 well plate; fluorescence was recorded every 1 °C from 25–95 °C on a Bio-Rad CFX Connect qPCR machine (Bio-Rad, Hercules, CA). The melting temperature for 50% of the probe (T_M) was determined with GraphPad Prism 5.0 using non-linear regression modeling (GraphPad Software, Inc., La Jolla, CA).

2.7. *Cellular Viability Assay*: Pancreatic cancer Panc-1 cells were obtained fresh from ATCC (Manassas, VA), and were maintained in Dulbecco's minimal essential medium supplemented with 10% fetal bovine serum and 1x penicillin/streptomycin solution at 37 °C, in a humidified atmosphere containing 5% CO₂, in exponential growth. For cellular viability assays, cells were seeded in 96-well plates at a concentration of 7.5×10^3 cells per well in 90 μ L of media, and allowed to attach overnight. The following day, a 10x stock plate of TMPyP2 or TMPyP4 were diluted from 5 mM over a 5-6 log range in 0.5 log steps was made and 10 μ L of this stock was added to the cell plate, in triplicate. The final high dose range was 0.08 – 500 μ M. Cell-free wells with the same dose-range were plated, and served as measurements of background

absorption. 48 hr later, 20 μ L of MTS + 5% PMS was added to each well and incubated for ~2 hr before the absorption at 490 nm was measured on a BioTek Synergy 2 plate reader (Winooski, VT) [32]. In parallel, media was removed from all wells and replaced with 100 μ L of fresh media before the addition of MTS as above. Background absorptions were subtracted, and data were normalized to control cells; IC₅₀ values were determined with GraphPad Prism software (San Diego, CA) using non-linear regression modeling.

2.8. qPCR: Panc-1 cells were plated in 6-well plates at a concentration of 2×10^5 cells per well in 1 mL of media, and were allowed to attach overnight. The following day, the media was changed to contain either vehicle, or 25 μ M TMPyP2 or TMPyP4. 48 hr later, RNA was harvested from the cells using the Thermo Scientific GeneJet RNA Purification kit (Fisher Scientific); yield and quality were determined with a NanoDrop 2000, and only samples with 260/230 values >2 were used for further analysis. cDNA was synthesized with the Bio-Rad iScript cDNA synthesis kit, and qPCR was run on a Bio-Rad CFX Connect real-time PCR detection system using TaqMan primers from ABI (kRAS:hs00364282_m1, GAPDH:hs99999905_m1). kRAS mRNA expression was normalized to GAPDH, and to untreated control by the $\Delta\Delta C_q$ method. Experiments were run in triplicate with internal technical duplicates; one-way ANOVA with post-hoc Tukey analysis was used to determine statistical significance.

2.9. Plasmid construction: A series of luciferase plasmids were constructed on the backbone of pGL4.17 plasmid (Promega). The promoter regions of interest, as denoted in **Figure 5**, were inserted between the Bgl I (kRAS-500 plasmid) or Nhe I (all kRAS-324 plasmids) and HIND III cut sites. kRAS-324 was constructed by Operon; the kRAS-500, 324 mt Near, and 324 mt Mid constructs were created in house according to previously published methods [33, 34] or by site-

directed mutagenesis with Q5 Site Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA), respectively. The MYC-promoter containing Del4 luciferase plasmid was obtained from Addgene (Cambridge, MA).

2.10. Transfection and Luciferase Assays: HEK-293 cells were maintained in Dulbecco's minimal essential medium supplemented with 10% fetal bovine serum and 1x penicillin/streptomycin solution at 37 °C, in a humidified atmosphere containing 5% CO₂, in exponential growth. Before transfection with luciferase plasmids, cells were seeded in six-well plates at a concentration of 2×10^5 cells per well, and allowed to attach overnight. Each well was co-transfected with the plasmid of interest (500 ng) and the reference renilla plasmid pRL-SV40 (125 ng) with FuGene HD (Promega) in a 3:1 ratio with cells in OptiMEM medium (Invitrogen; Grand Island, NY). An unmodified pGL4.17 plasmid (Empty Vector, EV) and a constitutively active SV40-driven pGL4.13 plasmid were used as control vectors. Transfections were maintained at 37 °C in a 5% CO₂ humidified environment overnight before media was replaced with DMEM ± TMPyP2 or TMPyP4 in dark conditions, as the compound may be activated to a photosensitizer. After 48 hrs of incubation, the expression of firefly, with respect to renilla, luciferase was determined with the Dual Luciferase Assay kit (Promega); light output was measured with a Lumat LB9507 luminometer. All experiments were performed with biological triplicate with internal technical duplicates. Statistical significance was determined by a one-way ANOVA with post-hoc Tukey post-hoc analysis, or a two-way ANOVA, as indicated in the text.

3. Results

3.1. New G4-forming regions within the kRAS promoter

The core promoter region of kRAS is encompassed in the region from -550 – +50 bp surrounding the TSS [9-11], and contains cis-regulatory elements [34] and silencing G4 elements [33]. Examination of the entire 500 bp core promoter revealed two previously undescribed putative G4-forming regions in addition to the previously described region [20, 21, 33](**Figure 1A**). Herein, we term the known 32-bp sequence, found -128 bp from the TSS, as the near region, forming G4_{near}, the next distal region from -174 to -226 as the mid region, forming G4_{mid}, and the furthest region from -238 to -273 as the far region, potentially forming G4_{far}. The mid region consists of 52 bases and seven runs of three or more continuous guanines with intervening loops of 2-7 nucleotides, and the far region consists of 35 bases and four runs of three or more guanines with intervening loops of 5-9 nucleotides.

G4 formation and overall stability of the near-, mid-, and far-regions was examined by CD, each in the absence and presence of intramolecular G4-stabilizing KCl alone or with 40% anhydrous acetonitrile (**Figure 1B**). As noted in previous publications, G4_{near} consists of a stable all-parallel structure as noted by the positive spectral maxima at 262 nm in the presence of KCl with or without ACN. In the presence of KCl, the thermal stability of this structure increased from 46 to 59 °C, and increased again to 69 °C with the addition of ACN. The mid region forms a mixed parallel and antiparallel structure (G4_{mid}) as noted with the positive spectral maxima present at 263 and 290 nm, respectively. Thermal stability increases from 56 °C in solution alone, to 82 and over 95 °C in the presence of 100 mM KCl without and with 40% ACN, respectively. The far-region does not form a strong G4 structure, and spectral maxima suggesting

parallel and antiparallel orientations are only notable in the presence of both KCl and ACN; these structures were too weak to have thermal stability determined.

G-to-T knockout mutations were introduced into each G4-forming region (Table 1), and their higher order DNA formations were examined by CD in the presence of 100 mM KCl (**Figure 1C**). As predicted the G-to-T mutations disrupted ($G4_{\text{near-ko}}$, mutation confirmed in by DMS footprinting below) or abrogated ($G4_{\text{mid-ko}}$ and $G4_{\text{far-ko}}$) G4 formations. These sequences were used in EMSA analysis below to indicate migration of the linear species.

Electromobility shift assay (EMSA) was utilized to examine the inter-, versus intra-, molecular G4 formations (**Figure 1D**). Sequences with G-to-T knockout mutations introduced and examined above were used as linear reference strands; dashed lines drawn horizontally indicate the migration pattern of linear DNA for each sequence. For both the near and the mid-regions, in the presence of 100 mM KCl alone a downward shift is evident, and the addition of 40% ACN makes the shift more marked. These faster migrations support the formation of smaller G4 structures under these conditions. The $G4_{\text{far}}$ sequence in the absence and presence of 100 mM KCl migrates lower than the linear knockout band, but there is no remarkable difference between those two. The addition of 40% ACN to this sequence retards migration of the DNA to be at or above the linear mark. Combined with CD data above, the $G4_{\text{far}}$ sequence is most likely forming a mixed parallel/antiparallel intermolecular structure under the combined conditions of cationic strength (KCl) and dehydration (ACN). The far region, by all data collected, is unlikely to form a significant intramolecular structure, and it was not examined any further.

3.2. Clarification of $G4_{\text{near}}$ isoforms from the complete 32-nucleotide region.

In an effort to refine the predominant G4 isoform forming from the sequence at -129 bp from the TSS, biophysical characterization was performed on the 32 nucleotide sequence 5'-AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG (**Figure 2A**). This was undertaken in order to clarify the array of previously reported structures that vary in their guanine runs and inclusion of flanking regions [20-23, 35]. In particular, DMS footprinting of the wild-type and two G-to-T mutant sequences demonstrated the major guanines used in the predominant isoform (**Figure 2B**). In the presence, as compared to the absence, of KCl, the DMS cleavage pattern for the induced G-quadruplex in the WT sequence revealed a strong utilization of guanine runs A – C, and equilibrating utilization of the guanines in run E with a preference for the central three. Run D appears to be hyper-reactive to DMS, as compared to the absence of KCl (**Figure 2B, left**). Due to the unique incorporation of a discontinuous guanine run (B), despite the presence of a possible fourth continuous guanine run (D), footprinting was also done on mtB and mtD sequences harboring G-to-T mutations within the runs indicated. When run B was mutated, the guanine protection pattern was disrupted and no higher order structure was noted; mutations of run D, however, maintained a pattern of guanine protection that was consistent with the WT sequence (**Figure 2B**). These data, cumulatively with the CD findings from **Figure 1B-C**, support modeling the predominant G4 formed from this 32-nucleotide region as an all parallel 1:1_k:1:12 loop isomer, with the 1_k indicating the kinked thymine in run B between guanines (**Figure 2C**). Collectively these data support that the kinked structure shown in the literature [23] is indeed the predominant isoform within the entire G4_{near}-forming-region.

3.3. Analysis of major G4's formed within the mid-region of the kRAS promoter

The 54-nucleotide mid-G4-forming region of the kRAS promoter contains seven unique runs of three or more continuous guanines, termed A–G in the 5'–3' direction (**Figure 3A**). DMS footprinting was performed on the entire wild-type G4_{mid} sequence without and with 100 mM KCl; histograms of the banding patterns were obtained with Image J software. It is clear that a number of higher order structures are forming in equilibrium, with at most only partial protection patterns clear in runs B, E and F (**Figure 3B**).

To aid in the preliminary description of the G4_{mid} structure, a series of single run knockout G-to-T mutations were made such that each run of continuous guanines was disrupted. These mutants were then examined by CD spectral and thermal analysis in conditions of varying cation and dehydration stability in order to probe the most relevant guanine runs involved in the predominant structure (**Figure 3C**). In particular, the sequences were studied in 10 mM KCl, 10 mM KCl + 20% acetonitrile, and 100 mM KCl (**Table 2**). All structures demonstrate mixed loop directionality by CD with maxima in the parallel (260–264 nm) and antiparallel (~290 nm) ranges (data not shown). Consistently, in all conditions, mutations of runs B, E, and F destabilized the G4 structure, as evidenced by a decrease in the melt temperatures or a marked change in the slope of the melting profile, which indicates a change in the predominant isoforms (**Figure 3C**). Melting profiles were examined in 100 mM KCl + 40% ACN conditions as well, but due to the extremely high thermal stability, T_{MS} were indeterminate; the trends in mutant effects on overall G4 stability were comparable with the lower KCl and ACN conditions (data not shown). Taking the CD and DMS data together, a major isoform noted under the various conditions has been proposed utilizing runs B, C, E and F as a triple stacked tetrad with intervening loops of lengths 2, 10, and 8 in the 5'–3' direction, respectively (**Figure 3D**).

3.4. Biological function of various G4s within the entire kRAS promoter

Previous literature has described kRAS promoter silencing in the presence of TMPyP4 [20, 33], which we used along with its positional isomer TMPyP2 to evaluate the biological function of the near- and mid-G4-forming regions of the kRAS promoter. The effect of each cationic porphyrin on the thermal stability of the G4_{near} and G4_{mid} structures was determined by FRET melt (**Figure 4A**). Consistently, 1 μ M TMPyP2 did not stabilize either structure, while 1 μ M TMPyP4 did. In particular, the T_M of G4_{near} was 34 °C alone and in the presence of TMPyP2, but increased to 65 °C in the presence of TMPyP4. Similarly, G4_{mid}'s T_M of 52 °C alone, or 46 °C in the presence of TMPyP2, was notably increased by TMPyP4 to 73 °C. While TMPyP4 does not show a preference to either G4 structure, it is a useful study tool to examine G4-mediated silencing of the kRAS promoter, with TMPyP2 as a negative control compound.

The effect of each compound on cellular viability and kRAS regulation was monitored in Panc-1 pancreatic cancer cells, which are often used in kRAS studies and particularly in previous G4 studies of the kRAS promoter [20, 21, 23, 36]. Changes in Panc-1 cellular viability were monitored with the MTS assay at concentrations up to 500 μ M incubated with the cells for 48 hr. As both of these compounds have a marked absorbance in the same range as formazin (e.g. 490 nm), the assay was performed both subtracting for the colorimetric contribution of matched doses of TMPyP2 and TMPyP4 (termed pre-wash) and after removing all media from the 96 well plates and replacing it in all wells with fresh media before the addition of MTS + 5% PMS (termed post-wash); background corrections were still made for the lingering colorimetric effects of the compounds. Once the corrections were made for compound, it is evident from the post-wash conditions that at the concentrations utilized in all future experiments, namely 25 and 50 μ M, there is no effect on Panc-1 cellular viability. Moreover, at concentrations up to 500 μ M,

TMPyP2 does not significantly impact Panc-1 cells, and TMPyP4 only decreases viability by approximately 25% (**Figure 4B**). Panc-1 cells were incubated with 25 μ M TMPyP2 or TMPyP4 for 48 hr, and changes in kRAS mRNA expression were determined by qPCR. TMPyP4 significantly ($p < 0.05$) decreased kRAS expression to $45.0 \pm 0.1\%$ of control expression, whereas TMPyP2 had no significant effect (**Figure 4C**). These findings support the role of G4-mediated silencing of kRAS expression.

In order to assess the contribution of G4_{near} and G4_{mid} to the silencing of kRAS transcription, a series of luciferase promoters were constructed (**Figure 5A**). In particular, regions of the kRAS promoter were inserted into the promoterless pGL4.17 firefly luciferase plasmid (empty vector = EV). Following previously published methods [34] we assembled a plasmid (kRAS-500) containing the promoter region of kRAS from -500 – +0 in relation to the TSS, which had been shown to contain silencing G4s [33]. In addition a plasmid was assembled containing -324 – +50 in relation to the TSS (kRAS-324) to contain the near, mid, and far-G4-forming region and to include the immediate post-transcriptional region, which had previously been shown to be critical for full promoter activity [9-11]. In addition, G-to-T mutations were introduced in either the near- (324 mt Near) or mid- (324 mt Mid) G4-forming region. These mutations matched those from the knockout oligonucleotides in **Figure 1C**. EV and an SV40-driven pGL4.13 firefly luciferase plasmid were used as control vectors for all experiments.

Basal expression from all plasmids was measured at 48 hr, and firefly expression was normalized to pRL-driven renilla luciferase expression (**Figure 5B**). SV40-promoter driven expression is almost 50-fold greater than activity from the promoterless EV (1.46 ± 0.08 versus 0.03 ± 0.01 , respectively). The kRAS promoter plasmids fell in between these two control plasmids, and did not vary significantly from each other. In particular, kRAS-500 basal

expression was 0.69 ± 0.11 and kRAS-324 was 0.94 ± 0.16 . Each of these plasmids was also transfected into HEK-293 cells, which were then exposed to 25 μ M TMPyP2 or TMPyP4 for 48 hr (**Figure 5C**). There was no significant change in promoter activity for any plasmid when exposed to TMPyP2, nor was there any effect of TMPyP4 on the EV or SV40 plasmids. However, both kRAS-500 and kRAS-324 demonstrated significantly lower promoter activity with TMPyP4 treatment; the decreased promoter activity the same for both plasmids. As all G4-silencing potential is housed within the 324 bp from the TSS, and the kRAS-324 plasmid contains the previously reported critical +50 bp region downstream of the TSS[9-11], further studies were done using this plasmid.

As described above, site-directed mutagenesis was used to introduce G-to-T mutations to the near- and mid-G4 forming regions such that higher order DNA structures could not form (as confirmed in **Figure 1C**), while minimizing disruptions to potential transcription factor binding sites. Basal expression was measured from each of the 324 plasmids after 48 hr of transfection, and there were no significant differences between any plasmids (**Figure 5D**). In particular, as compared to the kRAS-324's expression of 0.94 ± 0.16 , the expression for 324 mt Near and 324 mt Mid were 1.31 ± 0.23 and 0.96 ± 0.30 , respectively. Upon the addition of TMPyP2, as for all plasmids described above, no significant changes occurred with any plasmid with fold RLU of 0.99 ± 0.10 and 1.01 ± 0.08 for 324 mt Near and 324 mt Mid, respectively.

Upon exposure of the plasmids to 25 and 50 μ M TMPyP4, a number of significant effects were noted (**Figure 5E**). Specifically, higher concentrations of TMPyP4 led to an apparent decrease in luciferase activity or in luciferin light detection as evidenced by the significant change in fold RLU from both the EV and SV40 plasmids exposed to 50 μ M compound to 0.75 ± 0.05 and 0.72 ± 0.04 , respectively. The dose-response noted in EV and SV40 plasmids were

comparable, and were considered to account for background effects of TMPyP4. This non-specific effect is most likely due to dampening of the luciferin glow by the cationic compound, in agreement with a previous study showing the same decreased RLU but no transcriptional downregulation from the EV plasmid treated with 100 μ M TMPyP4, as measured by PCR [37]. Any significant changes in the extent of the dose response were compared to these background effects using a two-way ANOVA. Both the kRAS-324 and the 324 mt Near, but not the 324 mt Mid, plasmids had significantly ($p < 0.05$) different dose-responses than the control plasmids. At the highest dose, fold RLU was decreased to 0.55 ± 0.09 and 0.51 ± 0.15 for the kRAS-324 and 324 mt Near plasmids, respectively, and was 0.89 ± 0.14 for the 324 mt Mid plasmid. When the foldRLU is normalized to the EV and SV40 non-specific effects, a 26 and 30% decrease in RLU was observed for the kRAS-324 and 324 mt Near plasmids, respectively, while the 324 mt Mid plasmid displayed a 21% increase in RLU. For comparison, at the same concentration of TMPyP4, the MYC promoter-containing Del4 plasmid [38, 39] demonstrated a fold RLU of 0.56 ± 0.13 , for a 24% decrease as compared to the EV and SV40 plasmid non-specific effect (data not shown). From these data, we conclude that within the entire kRAS core promoter, the most critical higher order DNA structure related to transcriptional silencing is the G4_{mid} structure, as its abrogation disrupts TMPyP4-mediated promoter downregulation, and it is a new molecular structure for the development of targeted therapeutics.

4. Discussion

The current work has characterized an extensive guanine-rich region of the kRAS core promoter, extending out 500 base-pairs past the transcriptional start site. Within this region, of the three distinct putative G-quadruplex (G4)-forming areas, the most proximal and the medial regions, termed near and mid, respectively, formed inducible structures under a variety of buffer conditions, whereas the most distal region, termed far, did not. Much more critically, a series of experiments probing the potential biological role of G4 formation within cells highlighted the silencing function to be maintained predominantly in the mid region, versus in the previously described near structure. Initial probing indicates a great number of competing isoforms exist in equilibrium under ex vivo conditions, and suggest that a major $G4_{\text{mid}}$ structure is a tri-stacked mixed parallel and antiparallel isoform utilizing the second, third, fifth, and sixth runs of continuous guanines with loop lengths of up to 10 nucleotide. Extensive characterization work is ongoing to narrow the equilibrating isoforms under a variety of physiological conditions, including molecular crowding, dehydration, and torsional stress.

The core promoter region of kRAS is highly G/C-rich (~75%) and contains two nuclease hypersensitivity elements housing the described G4-forming regions [9-11]. Notably, the region +/- 50 bp surrounding the TSS is critical for directing and initiating transcription. The further upstream elements, including the DNase-sensitive G4-forming regions, are important for optimal kRAS expression and are capable of dampening promoter activity, but are not the main initiator regions for transcription [11]. Mutation of biologically active G4 structures within critical core promoters generally leads to a change in basal promoter activity [40, 41]. No such observation was made in the current study with mutation of either the near or the mid G4 structure in the

current study, which is in agreement with the previous description of the upstream promoter region.

The physical binding of transcription factors to the kRAS promoter region has not been mapped, making it difficult to assess the potential impact of mutating potential transcription factor binding sites. The core promoter +/- 50 surrounding the TSS contains one consensus Sp1 binding site, five E2F-1 sites (four of which are in the critical region from 0 - +50 bp), and consensus sites for a number of other factors, including WT1, GR-alpha, p53, STAT4, and NF-AT1/2. The near- and mid-G4-forming regions each contain a number of putative transcriptional regulator binding sites as well, including two for MAZ (one at the end of the near region and one in the midst of the mid region), three for Sp1 (one in the near region and two within the mid region), and several for p53, E2F-1, STAT4, WT1, NF-kB, and more. The G-to-T mutations introduced to the mid-G4 region disrupted predicted transcription factor binding sites for p53, E2F, E2F1, and WT1. Mutation of the near-G4 region interrupted putative binding sites for Sp1 (one of six within the kRAS-324 plasmid), p53 and E2F. It is possible that the disruption of transcription factor binding sites within these regions could lead us to underestimate the silencing potential of each G4-forming region, as the expected increase in promoter activity due to G4 mutation would be dampened by a loss of transcription factor binding. When the disruption of G4 formation is considered, however, along with the observed effects of TMPyP4 with the WT and each mutant plasmid, the likelihood of a significant silencing effect of the mid-G4 is high and that of the near-G4 is low.

The near region has been extensively explored by the Xodo group [20-23, 35], with variations in nucleotide inclusion, leading to evolving molecular models. These models varied from a di- to a tri-tetrad stack, altering loop lengths and directionality. Initially, 2 isoforms were

proposed – both parallel three-tetrad G4s with a kinked cytosine located between the first and second guanine runs [21]. In subsequent publications, the models were re-configured and inclusion of a thymine in run two was suggested [20]. The first model containing the possibility of a kinked thymine in the second guanine run was noted in 2009 with the inclusion of T/G (at the published positions 8/9) [23], and a more concrete description of the hypothesized isoform was offered in 2011 as a supplementary model [22]. Given the varying models and sequences, the present work included clarification of the particular guanines involved in structure formation from the near-G4-forming region in a more inclusive sequence, and confirmed the “kinked” structure as the predominant isoform.

Surprisingly, further examination into the biological function of this near sequence, in the context of the human core promoter, demonstrated a lack of silencing potential. This is in contrast to the conclusions of a 2006 study in which transfection of the human kRAS G4_{near} promoter sequence along with a CAT plasmid driven by the mouse kRAS promoter sequence indicated silencing potential [21]. The previous experiment examined the effect of transfecting wild-type and knockout mutant G4_{near} on the promoter activity from the mouse sequence (with 60-70% homology to the human sequence [11]), whereas the current study directly examines the human sequence with wild-type and specific G4_{near}, in addition to G4_{mid}, knockout mutations within the entire core kRAS promoter. While initial reports indicated a role of G4_{near} in silencing kRAS expression, as this was never directly examined within the context of the human sequence, the results presented herein are more accurate assessments of the biological function. These reports are not in disagreement, but are rather comparing different outcomes.

It is important to note that the data supporting G4-mediated-silencing of the kRAS promoter to be housed in the mid-G4-forming region does not negate other important findings

regarding G4 formation within the kRAS promoter. In particular, work with TMPyP4 suggested a decrease in kRAS expression within the first 500 bp from the TSS [33], which we confirmed and clarified further to exist within the first 324 bp. Additionally, there are a number of studies working with protein regulation of the kRAS promoter [20, 23, 42], and the use of structurally unrelated G4-decoys that have in vitro and in vivo efficacy [35, 36] highlighting the function of G4 formation. Rather, we hypothesize that the G4-mediated events noted in these other works in cells are mediated through the silencing G4 identified in the current works in the mid-G4-forming region. There are a number of similar sequences and consensus binding sites for the same transcription factors [9-11], further in support of this hypothesis.

kRAS is frequently mutated in multiple cancer types, especially in pancreatic cancer, making it a good anti-cancer target. The very low survival rate of pancreatic cancer patients clearly indicates that new and more efficacious treatments are needed and targeted kRAS downregulation holds a great deal of promise. Efforts have been made to develop clinical agents focused on kRAS such as targeting its membrane localization with farnesyltransferase inhibitors (FTIs), or its downstream effectors, such as Raf kinase, MEK, and mTOR [43, 44]. Unfortunately, none of these strategies have shown clinical efficacy. It is notable that specific inhibition of kRAS expression using antisense or siRNA oligonucleotides has shown promising preclinical activity, but their application in the clinic is hampered by difficulties in drug delivery [6, 7]. The characterization of a unique DNA structure, as described herein, allows for a new area of therapeutic research focused on small-molecule mediated downregulation of kRAS expression. The identification of small molecules that can interfere with *kRAS* transcription by stabilizing G-quadruplexes, will combine the best of all approaches – the specific downregulation of kRAS expression with the benefit of ease of delivery, as compared to

antisense oligos. Thus, such small molecules will have great potential in ultimately achieving clinical activity in patients whose tumors harbor this dysregulated oncogene.

Putative G-quadruplex forming regions of DNA are preferentially clustered ~1 Kb upstream of the transcriptional start site [45]. Interestingly, these sequences are found more frequently in oncogenic promoters [12], including some representatives of the hallmarks of cancer [46]. The varying loop lengths and tetrad composition of G-quadruplexes allows for specific targeting, similar to targeting a protein with a unique tertiary structure. The targeting of G-quadruplex secondary structures within oncogenic promoter regions has led to the development of two agents which advanced into clinical trials: the first-in-class small molecule Quarfloxin by Cylene Pharmaceuticals, which was halted at Phase II clinical development due to difficulties with delivery and excessive albumin binding, and Antisoma's G-rich phosphodiester oligonucleotide AS1411, a DNA aptamer with rare, but durable activity in renal cell carcinoma, with minimal associated toxicities [47].

We have identified and characterized a previously unexamined region of the kRAS promoter that is capable of forming a stable G4, of which stabilization by the cationic porphyrin TMPyP4 led to a significant decrease in promoter activity in an isolated plasmid system and in whole cells. TMPyP4 is a promiscuous G4-binding compound, although not of particular potency; its effect on kRAS promoter activity was consistent with the effect on the MYC promoter-containing Del4 plasmid, as noted in the results section. The MYC structure is the most well described promoter G4, with known silencer function [46, 48]. Thus, the similarity of observed changes in promoter activity from these two plasmids supports the promise of targeting the kRAS mid-G4 for clinical gain. The work with G4 decoys modulating the kRAS promoter and affecting tumor growth in vivo, albeit prescribed to a less significant region of the promoter

in the original publications, further highlights the potential G4-mediated regulation of kRAS transcription [35, 36]. This guanine-rich region of the kRAS promoter represents a highly valuable new molecular target for the development of small molecule therapeutics aimed at a number of cancers harboring mutant kRAS, most notably pancreatic cancer.

REFERENCES:

- [1] A.A. Adjei, Blocking oncogenic Ras signaling for cancer therapy, *Journal of the National Cancer Institute*, 93 (2001) 1062-1074.
- [2] B.B. Friday, A.A. Adjei, K-ras as a target for cancer therapy, *Biochimica et biophysica acta*, 1756 (2005) 127-144.
- [3] A. Young, J. Lyons, A.L. Miller, V.T. Phan, I.R. Alarcon, F. McCormick, Ras signaling and therapies, *Advances in cancer research*, 102 (2009) 1-17.
- [4] Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, RAS oncogenes: weaving a tumorigenic web, *Nature reviews*, 11 (2011) 761-774.
- [5] S.R. Hingorani, E.F. Petricoin, A. Maitra, V. Rajapakse, C. King, M.A. Jacobetz, S. Ross, T.P. Conrads, T.D. Veenstra, B.A. Hitt, Y. Kawaguchi, D. Johann, L.A. Liotta, H.C. Crawford, M.E. Putt, T. Jacks, C.V. Wright, R.H. Hruban, A.M. Lowy, D.A. Tuveson, Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse, *Cancer cell*, 4 (2003) 437-450.
- [6] A.M. Duursma, R. Agami, Ras interference as cancer therapy, *Seminars in cancer biology*, 13 (2003) 267-273.
- [7] E. Wickstrom, Oligonucleotide treatment of ras-induced tumors in nude mice, *Molecular biotechnology*, 18 (2001) 35-55.
- [8] K. Podsypanina, K. Politi, L.J. Beverly, H.E. Varmus, Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by Myc and mutant Kras, *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008) 5242-5247.
- [9] J. Jordano, M. Perucho, Chromatin structure of the promoter region of the human c-K-ras gene, *Nucleic acids research*, 14 (1986) 7361-7378.
- [10] J. Jordano, M. Perucho, Initial characterization of a potential transcriptional enhancer for the human c-K-ras gene, *Oncogene*, 2 (1988) 359-366.
- [11] F. Yamamoto, M. Perucho, Characterization of the human c-K-ras gene promoter, *Oncogene research*, 3 (1988) 125-130.
- [12] A. Verma, K. Halder, R. Halder, V.K. Yadav, P. Rawal, R.K. Thakur, F. Mohd, A. Sharma, S. Chowdhury, Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species, *J. Med. Chem.*, 51 (2008) 5641-5649.
- [13] J. Eddy, N. Maizels, Gene function correlates with potential for G4 DNA formation in the human genome, *Nucleic acids research*, 34 (2006) 3887-3896.
- [14] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat Chem*, 5 (2013) 182-186.
- [15] E.Y. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, G-quadruplex structures are stable and detectable in human genomic DNA, *Nature communications*, 4 (2013) 1796.
- [16] S. Muller, S. Kumari, R. Rodriguez, S. Balasubramanian, Small-molecule-mediated G-quadruplex isolation from human cells, *Nat Chem*, 2 (2010) 1095-1098.
- [17] V.S. Chambers, G. Marsico, J.M. Boutell, M. Di Antonio, G.P. Smith, S. Balasubramanian, High-throughput sequencing of DNA G-quadruplex structures in the human genome, *Nature biotechnology*, 33 (2015) 877-881.
- [18] C.K. Kwok, S. Balasubramanian, Targeted Detection of G-Quadruplexes in Cellular RNAs, *Angewandte Chemie*, 54 (2015) 6751-6754.

- [19] A. Shivalingam, M.A. Izquierdo, A.L. Marois, A. Vysniauskas, K. Suhling, M.K. Kuimova, R. Vilar, The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy, *Nature communications*, 6 (2015) 8178.
- [20] S. Cogoi, M. Paramasivam, B. Spolaore, L.E. Xodo, Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins, *Nucleic acids research*, 36 (2008) 3765-3780.
- [21] S. Cogoi, L.E. Xodo, G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription, *Nucleic acids research*, 34 (2006) 2536-2549.
- [22] M. Paramasivam, S. Cogoi, L.E. Xodo, Primer extension reactions as a tool to uncover folding motifs within complex G-rich sequences: analysis of the human KRAS NHE, *Chemical communications (Cambridge, England)*, 47 (2011) 4965-4967.
- [23] M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama, L.E. Xodo, Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription, *Nucleic acids research*, 37 (2009) 2841-2853.
- [24] R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek, V. Dötsch, The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions., *Nucleic acids research*, 39 (2011) 5768-5775.
- [25] M.C. Miller, R. Buscaglia, J.B. Chaires, A.N. Lane, J.O. Trent, Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3' Sequence of d(AG(3)(TTAG(3))(3)), *J Am Chem Soc*, (2010).
- [26] L. Petraccone, Higher-order quadruplex structures, *Topics in current chemistry*, 330 (2013) 23-46.
- [27] X. Qu, J.B. Chaires, Contrasting hydration changes for ethidium and daunomycin binding to DNA., *J. Amer. Chem. Soc.*, 121 (1999) 2649-2650.
- [28] X. Qu, J.B. Chaires, Hydration changes for DNA intercalation reactions., *J. Amer. Chem. Soc.*, 123 (2000) 1-7.
- [29] J.W.R. Schwabe, The role of water in protein–DNA interactions., *Curr. Opinion Struct. Biol.*, 7 (1997) 126-134.
- [30] R.J. DeSa, I.B. Matheson, A practical approach to interpretation of singular value decomposition results, *Methods in enzymology*, 384 (2004) 1-8.
- [31] A. De Cian, L. Guittat, M. Kaiser, B. Sacca, S. Amrane, A. Bourdoncle, P. Alberti, M.P. Teulade-Fichou, L. Lacroix, J.L. Mergny, Fluorescence-based melting assays for studying quadruplex ligands, *Methods (San Diego, Calif)*, 42 (2007) 183-195.
- [32] B.T. Mossman, In vitro approaches for determining mechanisms of toxicity and carcinogenicity by asbestos in the gastrointestinal and respiratory tracts, *Environmental health perspectives*, 53 (1983) 155-161.
- [33] J. Lavrado, H. Brito, P.M. Borralho, S.A. Ohnmacht, N.S. Kim, C. Leitao, S. Pisco, M. Gunaratnam, C.M. Rodrigues, R. Moreira, S. Neidle, A. Paulo, KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines, *Scientific reports*, 5 (2015) 9696.
- [34] I.S. Song, N.S. Oh, H.T. Kim, G.H. Ha, S.Y. Jeong, J.M. Kim, D.I. Kim, H.S. Yoo, C.H. Kim, N.S. Kim, Human ZNF312b promotes the progression of gastric cancer by transcriptional activation of the K-ras gene, *Cancer research*, 69 (2009) 3131-3139.
- [35] S. Cogoi, M. Paramasivam, V. Filichev, I. Geci, E.B. Pedersen, L.E. Xodo, Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-

- decoys with antiproliferative activity in pancreatic cancer cells, *Journal of medicinal chemistry*, 52 (2009) 564–568.
- [36] S. Cogoi, S. Zorzet, V. Rapozzi, I. Geci, E.B. Pedersen, L.E. Xodo, MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice, *Nucleic acids research*, 41 (2013) 4049–4064.
- [37] R.V. Brown, V.C. Gaerig, T. Simmons, T.A. Brooks, Helping Eve Overcome ADAM: G-Quadruplexes in the ADAM-15 Promoter as New Molecular Targets for Breast Cancer Therapeutics, *Molecules (Basel, Switzerland)*, 18 (2013) 15019–15034.
- [38] T.C. He, A.B. Sparks, C. Rago, H. Hermeking, L. Zawel, L.T. da Costa, P.J. Morin, B. Vogelstein, K.W. Kinzler, Identification of c-MYC as a target of the APC pathway, *Science*, 281 (1998) 1509–1512.
- [39] T. Lemarteleur, D. Gomez, R. Paterski, E. Mandine, P. Mailliet, J.F. Riou, Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase, *Biochemical and biophysical research communications*, 323 (2004) 802–808.
- [40] J. Seenisamy, E.M. Rezler, T.J. Powell, D. Tye, V. Gokhale, C.S. Joshi, A. Siddiqui-Jain, L.H. Hurley, The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4, *Journal of the American Chemical Society*, 126 (2004) 8702–8709.
- [41] X.-D. Wang, T.-M. Ou, Y.-J. Lu, Z. Li, Z. Xu, C. Xi, J.-H. Tan, S.-L. Huang, L.-K. An, D. Li, L.-Q. Gu, Z.-S. Huang, Turning off transcription of the *bcl-2* gene by stabilizing the *bcl-2* promoter quadruplex with quindoline derivatives, *Journal of medicinal chemistry*, 53 (2010) 4390–4398.
- [42] S. Cogoi, M. Paramasivam, A. Membrino, K.K. Yokoyama, L.E. Xodo, The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element, *The Journal of biological chemistry*, 285 (2010) 22003–22016.
- [43] J. Bendell, R.M. Goldberg, Targeted agents in the treatment of pancreatic cancer: history and lessons learned, *Current opinion in oncology*, 19 (2007) 390–395.
- [44] J.S. Macdonald, S. McCoy, R.P. Whitehead, S. Iqbal, J.L. Wade, 3rd, J.K. Giguere, J.L. Abbruzzese, A phase II study of farnesyl transferase inhibitor R115777 in pancreatic cancer: a Southwest oncology group (SWOG 9924) study, *Investigational new drugs*, 23 (2005) 485–487.
- [45] J.L. Huppert, A. Bugaut, S. Kumari, S. Balasubramanian, G-quadruplexes: the beginning and end of UTRs, *Nucleic acids research*, 36 (2008) 6260–6268.
- [46] T.A. Brooks, L.H. Hurley, The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics, *Nature reviews*, 9 (2009) 849–861.
- [47] J.E. Rosenberg, R.M. Bambury, E.M. Van Allen, H.A. Drabkin, P.N. Lara, Jr., A.L. Harzstark, N. Wagle, R.A. Figlin, G.W. Smith, L.A. Garraway, T. Choueiri, F. Erlandsson, D.A. Laber, A phase II trial of AS1411 (a novel nucleolin-targeted DNA aptamer) in metastatic renal cell carcinoma, *Investigational new drugs*, 32 (2014) 178–187.
- [48] T.A. Brooks, L.H. Hurley, Targeting MYC Expression through G-Quadruplexes, *Genes Cancer*, 1 (2010) 641–649.

Figure Legends

Figure 1. G4 formations in the extended *kRAS* promoter region. (A) The *kRAS* promoter shown from -324 to -39 bp relative to the transcriptional start site, contains three distinct guanine-rich regions (boxed), termed near, mid, and far in the 5'–3' direction. * denote G-to-T mutations within each region for knockout mutations shown in (C). (B) CD was used to determine G4 formation and stability within each of these G4-forming regions in the absence (solid line) or presence of 100 mM KCl alone (long dash line) or in the presence of 40% acetonitrile (ACN) (short dash line). Thermal stability from 20–100 °C are shown in the insets. (C) CD demonstrated a lack of inducible G4 formation in the presence of 100 mM KCl within each region with the selected G-to-T mutations in the knock-put (ko) sequences. (D) Electromobility shift assays were used to demonstrate the inter- versus intra-molecular G4 formations within each G4-forming region in the presence of 100 mM KCl with or without 40% ACN. A downward shifting of the DNA with the G4_{near} and G4_{mid} sequences, as compared to their linear ko sequences, indicates intramolecular structure formation, particularly evident in the presence of both cationic strength and dehydration. The G4_{far} sequence, in the absence and presence of 100 mM KCl, demonstrates a downward shift from the linear ko sequence, but there is no difference between these two solvent conditions. In contrast, in the presence of KCl and ACN, there is a lack of a downward shift and the presence of retarded migration, as compared to control and KCl alone, indicating the potential for intramolecular G4 species.

Figure 2. Predominant G4 isoforms formed within the near *kRAS* promoter. (A) The G4_{near} sequence contains four runs of continuous guanines – A, C-E – and one discontinuous run – B. The open circles indicate evidence of protection in the subsequent DMS footprinting, whereas

the black circles indicate no protection or even hypercleavage from DMS labeling. (B) DMS footprinting of the near promoter region of kRAS in the presence (+) or absence (-) of 100 mM KCl was performed with the wild-type (WT) or select G-to-T mutations in runs B (mtB) or D (mtD) sequences. In the absence of KCl, all guanines were labeled; upon the addition of 100 mM KCl, a protection pattern emerges in the WT sequence demonstrating a lack of DMS labeling of guanines in runs A-C and E. As this was an unusual pattern, runs B and D were individually mutated and those sequences were subject to DMS footprinting. G-to-T mutations of the discontinuous guanines in run B abrogated a clear protection pattern, whereas G-to-T mutations of the apparently non-incorporated run D were inconsequential to G4 formation. (C) Cumulatively, these data with the whole WT and select mt sequences were compiled to model a “kinked” thymine G4, including 5’-3’ loop sizes of 1, 1 kinked thymine, 1, and 12 bases. G = black circles, C = dark gray circles, T = light gray circles, A = white circles.

Figure 3. Predominant G4 isoforms formed within the mid-region of the kRAS promoter. (A) The G4_{mid} sequence contains seven runs of continuous guanines, A–G. The light gray circles indicate marked protection, the medium gray indicate partial protection, and the black circles indicate DMS-mediated piperidine cleavage. The open circles indicate evidence of protection in the subsequent DMS footprinting, whereas the black circles indicate no protection or even hypercleavage from DMS labeling. (B) DMS footprinting of the mid promoter region of kRAS in the presence (+) or absence (-) of 100 mM KCl was performed with the wild-type (WT) sequence. Images obtained from the top and bottom portions of the sequencing gel (boxed individually) were aligned, Image J software was used to graph and align histograms of the guanine cleavage pattern (right, no KCl = black line; 100 mM KCl = red line). (C) Thermal

stability of a series of G-to-T mutants interrupting runs A–G, individually, of the G4_{mid} sequence were studied in the presence of 10 mM KCl alone (top), in the presence of 20% ACN (middle), or with 100 mM KCl (bottom). (D) Cumulatively, these data were used to predict a G4 isoform formation of a tri-stacked structure incorporating runs B, C, E and F with intervening loops of 2, 10, and 8 bases in the 5'–3' direction. G = black circles, C = dark gray circles, T = light gray circles, A = white circles.

Figure 4. Effect of G4 stabilization on pancreatic cancer cell viability and kRAS mRNA expression. (A) FRET Melt demonstrated a marked stabilization of both the G4_{near} and the G4_{mid} formations by the pan-G4-stabilizing TMPyP4 (1 μ M), but not the inactive positional isomer TMPyP2 (1 μ M). (B) The effect of each compound on pancreatic cancer cell, Panc-1, viability was measured at 48 hr. MTS assays were done with plates where the subtracted background included compounds (pre-wash) and a series in which the media with and without compound was removed and was replaced with fresh media only, and subsequently included in the background normalization (post-wash). Due to the contribution of compound to the absorbance at 490, this washing even had a marked effect on the fold cell viability. Post-washing, neither TMPyP2 nor TMPyP4, at doses up to 500 μ M, decreased cellular viability by 50%. Doses used in subsequent cellular assays are denoted by vertical dotted lines at 25 and 50 μ M. (C) Panc-1 cells were incubated with 25 μ M of either TMPyP2 or TMPyP4 for 48 hours; subsequently kRAS mRNA, as normalized to GAPDH, was measured. TMPyP4, but not TMPyP2, significantly (* p <0.05) decreased transcription and support the presence of G4-mediated silencing of kRAS expression. Experiments were done in a minimum of triplicate; One-way ANOVA with Tukey post-hoc analysis was used to determine significance in (C).

Figure 4.

Figure 5. Isolating silencing G4 formations within the *kRAS* promoter. (A) A series of luciferase plasmids were constructed from the promoterless empty vector (EV) pGL4.17 backbone to localize the silencing G4s within the *kRAS* promoter. These included *kRAS*-500 (-500 – +0, relative to the transcriptional start site (TSS)), and *kRAS*-324 (-324 – +50 relative to the TSS). Site-directed mutagenesis of *kRAS*-324 was used to introduce G-to-T mutations (approximate location indicated with white x) that abrogate G4_{near} (324 mut Near) or G4_{mid} (324 mt Mid) formation. (B) Basal expression of the non-mutated plasmids was examined in transiently transfected HEK-293 cells for 48 hr and normalized to renilla expression in co-transfection assays. Promoterless (EV) and constitutively active (SV40) plasmids were included as comparison. There was no significant difference between the *kRAS* promoter plasmids. * $p < 0.05$ as compared to EV, § $p < 0.05$ as compared to SV40 plasmid, as determined by one-way ANOVA. (C) The effects of HEK-293 treatment (48 hr) with 25 μ M TMPyP2 and TMPyP4 on promoter activity was examined in the EV, SV40 and wild-type *kRAS* promoter containing plasmids *kRAS*-500 and *kRAS*-324. TMPyP4 was equally significant ([#] $p < 0.05$ as compared to untreated control, per plasmid, as determined by one-way ANOVA) in lowering promoter activity in the two *kRAS* promoter plasmids, whereas it was inactive in the non-G4-containing EV and SV40 plasmids at that concentration. TMPyP2 was also inactive in all plasmids. (D) 48 hr Basal expression was examined from the 324 series of *kRAS* promoter plasmids, including wild-type *kRAS*-324, and select G4-knockout mutants 324 mt Near and 324 mt Mid. No change in basal expression was noted between any of these plasmids (E) A TMPyP4 dose response (25 and 50 μ M) was performed in the wild-type *kRAS*-324 plasmid and its two generated mutant plasmids 324 mt Near and 324 mt Mid. The dose response was performed in the non-G4-containing EV

and SV40 plasmids. The colorimetric contribution of 50 μ M TMPyP4 is evident in the EV and SV40 containing plasmids. In all plasmids where a dose-response was evident, the magnitude of that response was examined by a two-way ANOVA. Both kRAS-324 and 324 mt Near maintained significant ($\hat{p}<0.05$ for dose response, as compared to EV and SV40 effects) G4-mediated silencing of promoter activity, whereas 324 mt Mid did not. These findings indicate that G4-mediated silencing is contained within the mid-guanine-rich region of the kRAS promoter. All experiments were performed in a minimum of triplicate.

Figure 1

Figure 2

B

C

Figure 3

Figure 4

A

B

C

Figure 5

A

B

C

D

E

Table 1. Oligonucleotide sequences.

Name	5'-3' sequence
PSA primer	TCGACTCTAAGCAAATGCGT
G4 _{near}	AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG
G4 _{near} DMS	TTTTTTTAGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGGTTTTTTT
G4 _{near} FRET	[6-FAM]-AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG-[TAMRA]
G4 _{mid}	CGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGC
G4 _{mid} DMS	[6-FAM]-TTTTTTTCGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGCTTTTT
G4 _{mid} FRET	[6-FAM]-CGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGC-[TAMRA]
G4 _{far}	AAGGGGTGGCTGGGGCGGTCTAGGGTGGCGAGCCGGGCC
mut A	CGTTGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGC
mut B	CGGGGAGAAGGAGGTGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGC
mut C	CGGGGAGAAGGAGGGGGCCGCTGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGGGC
mut D	CGGGGAGAAGGAGGGGGCCGGGCCGCTGCCGGCGGGGGAGGAGCGGGGGCCGGGC
mut E	CGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGTGGAGGAGCGGGGGCCGGGC
mut F	CGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGTGGCCGGGC
mut G	CGGGGAGAAGGAGGGGGCCGGGCCGGGCCGGCGGGGGAGGAGCGGGGGCCGTGC

	Wild-type	Mut A	Mut B	Mut C	Mut D	Mut E	Mut F	Mut G
10 mM KCl	70	65	50	63	65	63	53	65
10 mM KCl + 20% ACN	73	70	64	71	71	64	64	71
100 mM KCl	81	78	70	89	82	84	63	87

Table 2. Melting temperatures for G4_{mid} sequences.