

AWARD NUMBER: W81XWH-14-1-0219

TITLE: Novel Molecular Targets for kRAS Downregulation: Promoter G-Quadruplexes

PRINCIPAL INVESTIGATOR: Dr. Tracy Brooks

CONTRACTING ORGANIZATION: University of Mississippi  
University, MS 38677-9704

REPORT DATE: September 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release, distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

*Form Approved*  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE</b> September 2015		<b>2. REPORT TYPE</b> Annual		<b>3. DATES COVERED</b> 15Aug2014 - 14Aug2015	
<b>4. TITLE AND SUBTITLE</b>  Novel Molecular Targets for kRAS Downregulation: Promoter G-Quadruplexes				<b>5a. CONTRACT NUMBER</b> W81XWH-14-1-0219	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  Tracy Brooks  E-Mail: tabrooks@olemiss.edu				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  University of Mississippi University, MS 38677-9704				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The overall objective of this project is to characterize the formation and regulation of newly identified, biologically active, higher order DNA structures in the promoter of kRAS, which is a signaling molecule that has been shown to have aberrant activity in over 90% of pancreatic cancers. The higher order DNA structure under examination is a G-quadruplex, capable of forming in guanine-rich DNA regions found in regulatory regions of DNA, such as telomeres, centromeres, 5' UTR's and promoters. Generally these structures function to silence transcription, although each promoter structure requires individual examination for a functional determination. Within the kRAS promoter lies an extensive guanine-rich region of DNA with three separate putative G-quadruplex forming regions, which we have termed near, mid, and far in relation to their proximity to the transcriptional start site. The near region has been previously described, but we have shown it to be biologically inert, in contrast to the more distal mid-G4-forming region. Progress within the funded project, to date, has focused on this newly identified DNA region, and has characterized predominating isoforms under varied physiological conditions (Aim 1). Additionally, we have examined the transcriptional regulation of the kRAS promoter, with particular attention to the dynamic structures formed within the mid-G4-forming region, by the transcriptional regulators Sp1, MAZ, and p53 (both wild-type and mutant). Cumulatively, findings from this proposal will have described a novel molecular target with detailed structural and regulatory information, enabling a concentrated drug discovery program with marked promise for new pancreatic cancer therapeutics.					
<b>15. SUBJECT TERMS</b>  Nothing listed					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER</b> (include area code)
Unclassified	Unclassified	Unclassified	Unclassified	21	

**Table of Contents**

	<b><u>Page</u></b>
<b>Introduction.....</b>	<b>1</b>
<b>Body.....</b>	<b>1</b>
<b>Key Research Accomplishments.....</b>	<b>1</b>
<b>Reportable Outcomes.....</b>	
<b>Conclusion.....</b>	
<b>References.....</b>	<b>B#5</b>
<b>Appendices.....</b>	<b>B#5</b>

## 1. Introduction

The overall objective of this project is to characterize the formation and regulation of newly identified, biologically active, higher order DNA structures in the promoter of *kRAS*, which is a signaling molecule that has been shown to have aberrant activity in over 90% of pancreatic cancers. The higher order DNA structure under examination is a G-quadruplex, capable of forming in guanine-rich DNA regions found in regulatory regions of DNA, such as telomeres, centromeres, 5' UTR's and promoters. Generally these structures function to silence transcription, although each promoter structure requires individual examination for a functional determination. Within the *kRAS* promoter lies an extensive guanine-rich region of DNA with three separate putative G-quadruplex forming regions, which we have termed near, mid, and far in relation to their proximity to the transcriptional start site. The near region has been previously described, but we have shown it to be biologically inert, in contrast to the more distal mid-G4-forming region. Progress within the funded project, to date, has focused on this newly identified DNA region, and has characterized predominating isoforms under varied physiological conditions (Aim 1). Additionally, we have examined the transcriptional regulation of the *kRAS* promoter, with particular attention to the dynamic structures formed within the mid-G4-forming region, by the transcriptional regulators Sp1, MAZ, and p53 (both wild-type and mutant). Cumulatively, findings from this proposal will have described a novel molecular target with detailed structural and regulatory information, enabling a concentrated drug discovery program with marked promise for new pancreatic cancer therapeutics.

## 2. Keywords:

G-quadruplex, *kRAS*, Sp1, p53, MAZ, transcriptional control

## 3. Accomplishments:

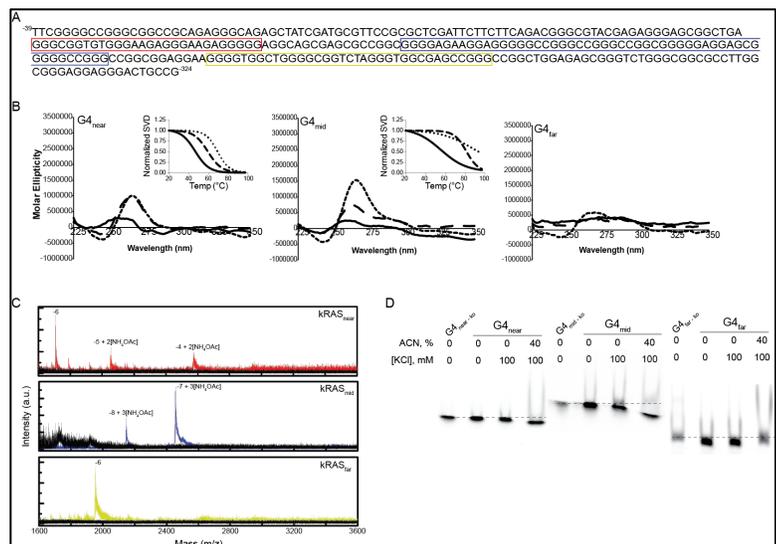
The **major goals** of this project were to (1) *Determine the predominant G4 structures within the mid- and far-regions of the kRAS core promoter*, and to (2) *Establish dynamic G4 regulation by candidate endogenous proteins*. Marked progress has been made in each of these two goals as follows:

### What was accomplished under these goals?

(1) Determine the predominant G4 structures within the mid- and far-regions of the *kRAS* core promoter.

While the near-G4-forming region had been described in the literature previous to the current project, we were the first to examine higher order formations within the mid- and far-G4-forming regions (Figure 1). Through electronic circular dichroism (ECD), electron-spray ionization – mass spectroscopy (ESI-MS), and EMSA, we were able to confidently conclude that a major inducible G-quadruplex does not form in the far-region, while a strong structure can be identified within the mid-region.

ESI-MS can be used to identify the number of ions associated with particular structures;  $\text{NH}_4\text{OAc}$  is more ideal for these studies, as compared to KCl, due to its more easily distinguished weight. The data highlights peaks without associated  $\text{NH}_4\text{OAc}$  in the near and far regions (both charged -6),



**Figure 1.** (A) *kRAS* promoter sequence with the near (red), mid (blue), and far (yellow) G-quadruplex-forming regions outlined is capable of forming (B) inducible G-quadruplexes as noted by increased maxima measured by ECD in the presence of 100 mM KCl alone (long dash) or in the presence of 40% acetonitrile (ACN, short dash), as compared to alone (solid line). The far-region only demonstrates inducible structures in the combined presence of cationic and dehydrating stressors. (C) The number of stacked tetrads was confirmed by ESI-MS, and three tetrads were noted for the near structure, while four were noted for the mid structure. No marked formations were identified in the far-region. (D) EMSA evaluation under cationic (KCl) and dehydrating (ACN) conditions demonstrated downward migrating structures in the near and mid-G-quadruplex-forming regions, but not in the far region.

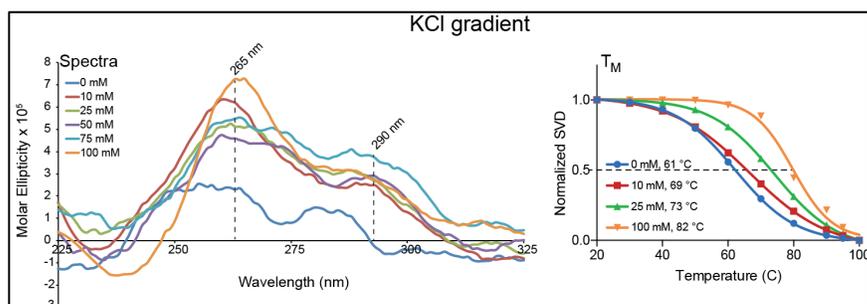
representing linear DNA. In the near and mid kRAS spectra, however, charged species with incorporated  $\text{NH}_4\text{OAc}$  is noted. These  $\text{NH}_4\text{OAc}$  incorporate between the stacked tetrads, so the number of tetrads can be calculated as  $[\text{NH}_4\text{OAc}]_{n+1}$ . Thus, the two peaks within the near region harboring 2  $\text{NH}_4\text{OAc}$  are two distinctly folded G4s, each with three stacked tetrads, and the two peaks in the mid region represent two unique G4 isoforms with four tetrads each (Figure 1C).

In data not shown, no changes in ECD of the near, mid, or far region were noted with the addition of up to 100 mM NaCl. This finding is supportive of the lack of intermolecular G4 structures, which are more easily stabilized by the smaller monovalent cation, in contrast to KCl.  $\text{NH}_4\text{OAc}$  was not examined in further studies as, although it was ideal for ESI-MS, data from the footprinting studies described below made it evident that the larger cation stabilized a non-dominant isoform. For this reason, the ESI-MS was removed from the published studies, although it is shown here as a portion of the work completed.

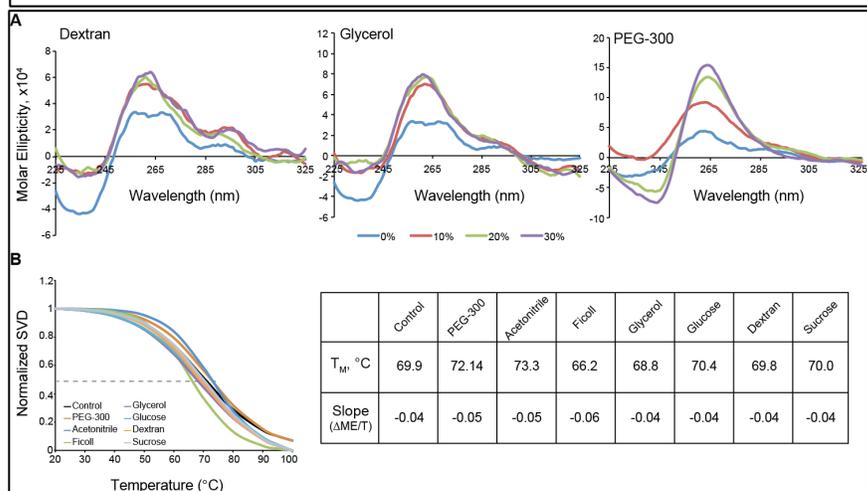
The more narrow focus of the further studies to the mid region only were further supported by plasmid studies demonstrating that knock-out mutations of the mid-region *only* were necessary and sufficient to abrogate G-quadruplex-mediated transcriptional silencing of the kRAS promoter (Morgan, et al, BBA-GRM, 2015).

The mid-G4-forming region is complex in its higher order structure formations, due to the presence of seven distinct runs of three or more continuous guanines (see Figure 1A in blue). Primers have been optimized for plasmid footprinting, and we are currently working to expand our footprinting studies to include the presence of nucleoplasm as a physiologic condition, as well as to use ligation-mediate PCR to footprint nascent DNA. To date, we have studied the G4 formation in cationic (KCl; **Figure 2**), and chemical (ACN, PEG, glucose, sucrose, glycerol, dextran sulfate, and Ficoll70, to cover dehydrating and crowding conditions alike, **Figure 3**). Of these, ACN and PEG demonstrate the most marked structural and thermal stabilization. Subsequent to these findings, PEG has been shown to have direct interactions with secondary DNA structures, rather than a general influence on the solvent, and further studies were done with the dehydration of ACN.

Chemical footprinting was performed for the mid-G-quadruplex-forming region, under varying co-solvent conditions (**Figure 4A-C**). Due to the complexity of the region, a number of isoforms exist in equilibrium, and a partial protection



**Figure 2.** Spectra (left) and Thermal evaluation (right) of the mid-G-quadruplex with increasing  $[\text{KCl}]$ , mM. 10 mM was noted to have a marked shift in both spectra and  $T_M$ , and was used in subsequent studies as a baseline for stimulating G4 formation; 100 mM demonstrated the most marked spectral and thermal changes, representing a physiologically relevant nuclear concentration.



**Figure 3.** Representative co-solvent spectra (A) from 0-30% w/v. These data were used to define the lowest % of co-solvent capable of markedly modulating G4 formation within the mid-forming region. At these lowest shifting concentrations + 10mM KCl, the effect on thermal stability (B) of the mid-G4-region was evaluated. Ficoll decreased the thermal stability of the mid-region with an increase in slope indicative of a deconvolution of multiple competing isoforms, while Glycerol, Glucose, Dextran and Sucrose did not change the fundamental G4-forming properties. Both dehydration with Acetonitrile and crowding with PEG-300 increased the  $T_M$  of the mid region while narrowing the landscape of G4 isoforms, as noted with a change in slope. Data not shown are with each co-solvent at 40%; patterns are similar and samples with either ACN or PEG-300 are unable to be melted at 100 °C.

pattern is evident. In order to assess the protection pattern, the optical density (OD) of the guanine bands were determined using Image J (NIH, Bethesda, MD) software, and the density was normalized to the proximal discontinuous guanines (e.g. lone or duplicate G's) in order to account for a variation of band intensity across the gel. Normalized OD is graphed in (Figure 4C), and was used to highlight the levels of partial protection (circles in Figure 4A). Specifically, dark circles represent unprotected guanines, dark gray circles highlight <50% protection, and light gray circles correlate with  $\geq 50\%$  protection. As an example, the protection in run B is denoted by \* for the region with  $\geq 50\%$  protection, as compared to #.

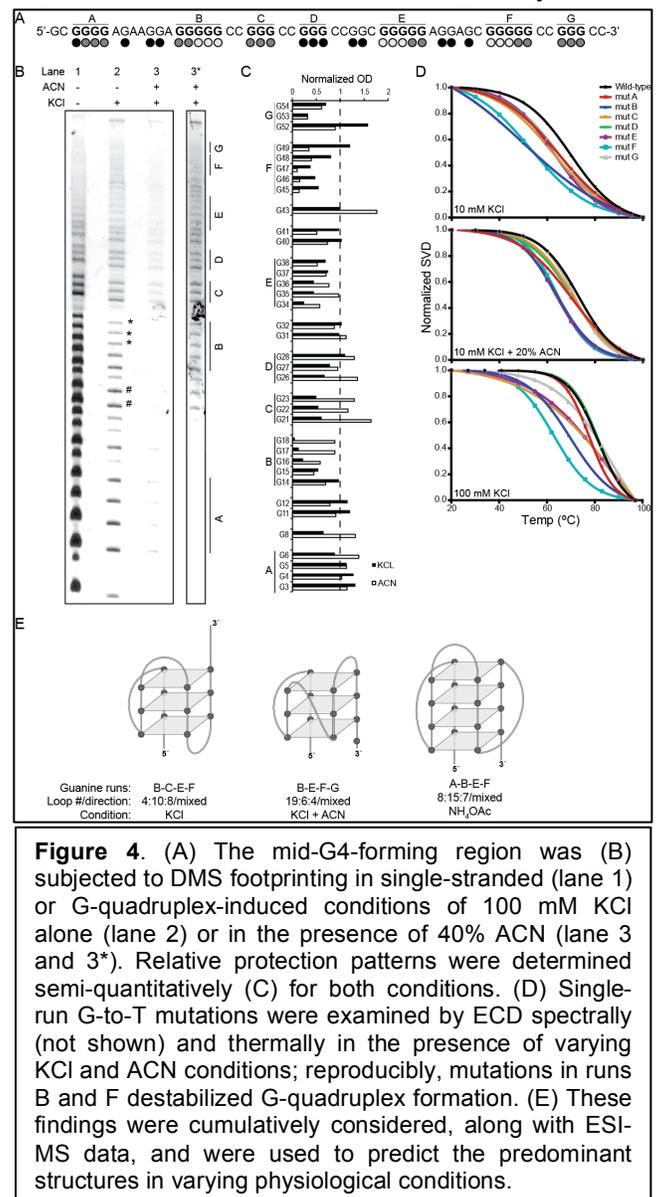
A number of point mutations were designed and examined to knock out individual guanine runs, and the impact of those mutations on ECD spectra (not shown) and thermal stability (Figure 4D) were examined, again under varying co-solvent conditions. The [KCl] varied from the minimal amount capable of stimulating G4 formation (10 mM) to the highest physiological concentration (100 mM). Based on the findings described above (Figure 3B), 20% ACN was used to dehydrate the samples. From these cumulative works, in addition to the ESI-MS shown in Figure 1, G-quadruplex models were predicted (Figure 4E). The most physiologically relevant structure is believed to be that of KCl + ACN, and it is currently being compared to findings with nucleoplasm in single-stranded and supercoiled DNA.

(2) Establish dynamic G4 regulation by candidate endogenous proteins.

There are literature reports on proteins binding the near G-quadruplex-forming region of the kRAS promoter, namely MAZ, and regulating transcription; through this aim we intend to describe the regulation of the mid-region. Through mapping transcription factor binding sites, and with the mindset of proteins with putative binding to higher order DNA structures, we proposed examining the binding of Sp1, MAZ, WT1 (not yet unway) and both wild-type (WT) and R273H mutated p53.

Our first experiments examined the regulation of kRAS transcription by the aforementioned putative regulatory proteins. To do so we followed a two-prong approach using the isolated promoter region in the luciferase plasmid, and monitoring endogenous regulation in a panel of pancreatic cancer cells (BxPc3 with WT kRAS, Panc-1 and Capan-1 each with MT kRAS and with moderate and high oncogene addiction, respectively) (Figure 4). Intriguingly, through HEK-293 transient co-transfections of protein overexpression plasmids and the kRAS promoter luciferase construct, we found that both MAZ and Sp1 facilitated a decrease in promoter activity, while both WT and MT p53 demonstrated no effect (Figure 4A). When MAZ and Sp1 were combined, the silencing effect was cancelled out and cumulatively there was no change in promoter activity (Figure 4B).

The effect of MAZ overexpression in endogenous regulation of kRAS transcription was studied in BxPc3, Panc-1 and Capan-1 pancreatic cancer cells. There were previous studies indicating that if Panc-1 cells were transfected with 500 ng of this CMV-MAZ plasmid, that MAZ transcription would increase 5-fold and kRAS transcription would mirror with a ~3.5-fold increase. However, despite our use of a panel of cell lines



and plasmid transfections up to 1000 ng with corresponding increased MAZ mRNA up to 6000-fold, we were unable to observe any modulation of kRAS expression (Figure 4C). This result was so unexpected, that we chose to obtain fresh new cells from ATCC to repeat the studies, in case genetic drift had occurred, but the findings did not change. The concentration of MAZ overexpression plasmid transfected into these cells was increased to 2000 ng, with which mRNA increased 3000-14000-fold; at these exorbitant expression levels, kRAS mRNA was seen to change 7-25-fold (data not shown).

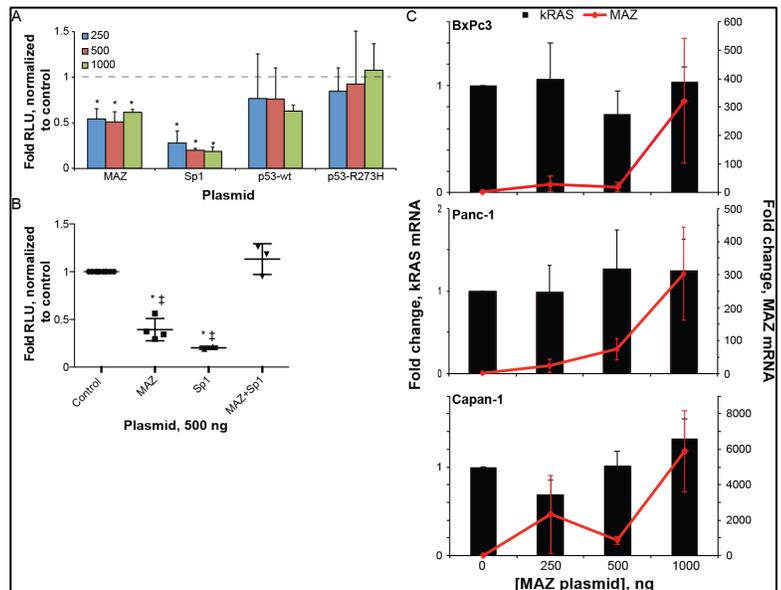
Binding interactions between SP1 or MAZ and the near- or mid-G4-quadruplex was monitored by electrophoretic mobility shift assay (EMSA). 1  $\mu$ M FAM-labeled DNA was induced into a higher order structure with 100 mM KCl, and was incubated with the denoted concentration of protein in binding buffer for 30 minutes prior to electrophoresis (Figure 5). Neither protein demonstrated supershifts in DNA at concentrations up to 4  $\mu$ g. Coomassie staining of the gels highlighted the MAZ or SP1 protein to be at a higher migration level and confirmed that at these concentrations neither protein is bound to the near- or mid-G-quadruplex, as formed in ssDNA. Without a significant mobility shift, we are unable to define binding affinities and coefficients. We are currently concentrating on identifying other proteins capable of interacting with the region of interest, as described in the plans below for Aim 2.

### What opportunities for training and professional development has the project provided?

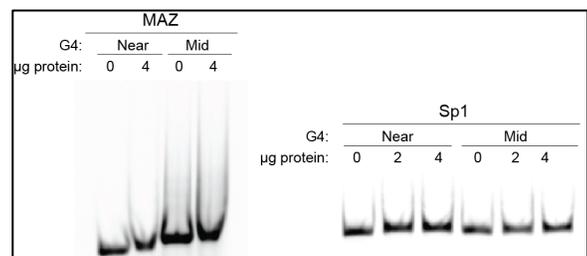
As a Career Development grant, training and development is a critical component of the aims. To this end, Dr. Brooks and her two graduate students working on this project were afforded the opportunity to attend the annual national meeting of the American Association of Cancer Research in Philadelphia, PA (April, 2015). At this meeting there was a large portion of time and resources dedicated to focusing on kRAS-related research, and the entire research team attended and extensively networked and expanded their development through these sessions. In addition, Dr. Brooks served as a grant reviewer for the DoD BCRP funding mechanism, which is a remarkable mechanism for professional development in developing grant writing and reviewing skills.

### How were the results disseminated to communities of interest?

For her outreach efforts, Dr. Brooks was invited to be the keynote speaker at the first annual Pancreatic Cancer Action Network of Mississippi "Light the Night Purple" event in March of 2015. She also serves as the Faculty Advisor to the University of Mississippi's American Cancer Society's Relay for Life Committee.



**Figure 4.** (A) HEK-293 cells were transfected with a luciferase plasmid driven by the core kRAS promoter and varying doses of protein overexpression plasmids. After 48 hours, cells were lysed and luciferase activity was examined. Sp1 and MAZ significantly decreased promoter activity, whereas neither form of p53 had an effect. (B) As MAZ and Sp1 bind similar consensus sequences and are modulated by the same upstream proteins (CK1), their combined effect on promoter activity was monitored with 500 ng of overexpression plasmid. Despite each transcription factor individually silencing promoter function, in combination there was no effect. (C) The effect of MAZ overexpression on endogenous regulation of kRAS transcription was monitored in pancreatic cancer cells BxPc-3, Panc-1, and Capan-1. Despite MAZ mRNA expression increasing up to 6000-fold, there was no change in kRAS expression.  $n \geq 3$ ; \* $p < 0.05$  as compared to control and ‡  $p < 0.05$  as compared to MAZ+Sp1.



**Figure 5.** The binding of MAZ and Sp1 to FAM-labeled Near- or Mid-G-quadruplexes (as induced by 100 mM KCl) were examined. Neither protein, at concentrations up to 4  $\mu$ g, demonstrated a supershift of the DNA, indicating there is no protein:DNA interaction.

In addition to the manuscript described in "Products" below, several oral and poster presentations have been given by graduate students on the projects described. These are:

- Morgan, R; Rahman, KM; **Brooks, TA**. Structure Elucidation of G-Quadruplex within the mid-region of the kRAS Promoter and Identification of Stabilizing Small Molecules as Promising Transcriptional Silencers. Proceedings of the American Association of Cancer Research (AACR), April 2015. Abstract 1245.
- Batra, H; **Brooks, TA**. The Effect of the transcription factor MAZ on kRAS transcription: a role for the G-quadruplex. Proceedings of the American Association of Cancer Research (AACR), April 2015. Abstract 2137.
- Morgan, R\*; Rahman, KM; **Brooks, TA**. Structure elucidation of G-quadruplex within the mid-region of the kRAS promoter and identification of stabilizing small molecules as promising transcriptional silencers. Mississippi Academy of Sciences Annual Meeting, Hattiesburg, MS. February 2015. \*1<sup>st</sup> place poster winner\*
- Batra, H; **Brooks, TA**. The Effect of the Transcription Factor MAZ on kRAS Expression: A Role in Pancreatic Cancer. ACS Drug Discovery and Development Colloquium, Little Rock, AR. 2014.
- Backus, K; **Brooks, TA**. A Novel Role for p53 in kRAS Binding in Pancreatic Cells. ACS Drug Discovery and Development Colloquium, Little Rock, AR. 2014.

#### **Plans during the next reporting period.**

*Aim 1:* Over the next year, we plan to continue our work isolating the predominant G-quadruplexes within the mid-region of the kRAS promoter that is responsible for transcriptional silencing. To do so, we will continue our efforts to chemically footprint the region in ssDNA form in the presence of nucleoplasm, as well as in combinations of individual co-solvents as described above. The overall effort to minimize co-solvent conditions is to enable future compound screening efforts in a large scale, for which nucleoplasm is not well suited. With the primers for plasmid footprinting optimized, we will monitor higher order DNA structure as a function of physical (e.g. torsional) stress combined with either nucleoplasm or co-solvents with the ultimate goal being chromosomal footprinting by LM-PCR. This last assay requires marked optimization at each primer level, and the plasmid footprinting is the first step in the continuum of experiments to reach that goal. Finally, we will initiate studies on isolated G-to-T mutations both in the plasmid and in the ssDNA form to isolate individual G-quadruplexes and monitor both structural formations and physiological function via promoter activity.

*Aim 2:* The findings with protein regulation described above were not as anticipated. Namely, the lack of effect of MAZ on the regulation of kRAS transcription was not expected. We will continue our efforts to examine the role of Sp1 and p53 in the endogenous regulation of kRAS mRNA expression in the panel of pancreatic cancer cells lines and will initiate the WT1 studies. Rather than working with the isolated regions of ssDNA, we will use our PCR amplified entire kRAS core promoter region (or various truncated portions containing different guanine-rich regions) to examine the global binding of MAZ, Sp1, WT1, and p53. Should our findings continue to indicate that these prospective proteins are not critically involved in kRAS regulation, we will follow our alternative strategy from the grant application and proceed with subtractive DNA columns using the PCR amplified core promoter, followed by the individual regions (mid, and near) and topologies (dsDNA, ssDNA, and G4-DNA), with Panc-1 nuclear extract. Bound proteins will be isolated and identified through LC-MS/MS.

#### **4. Impact:**

##### **Impact on the development of the G-quadruplex field.**

The studies presented do contrast the findings of some literature reports. Namely, they highlight the importance of examining the entirety of a promoter region when examining transcriptional regulation by higher order DNA structures (e.g. looking at near, mid, and far-guanine rich regions, rather than only that in proximity to the transcriptional start site). While the near-G-quadruplex has been reported, through extensive promoter mapping of not only G-quadruplex formation, but also of *function* in the context of the entire kRAS core promoter, we were able to clarify that any silencing potential is held in the mid-region.

This has a large impact on other groups that are pursuing drug discovery efforts focused on the near-region, which to date have been unsuccessful. The field will shift to pursue the mid-G-quadruplex in the near future.

**Impact on other disciplines.**

Nothing to report.

**Impact on technology transfer.**

The plasmids made and utilized through this project have been requested by other universities so that they may optimize their compound screening efforts. MTA's are being drafted to share our resources, and we intend to deposit our major plasmids with AddGene for global availability.

**Impact to society and technology.**

Nothing to report.

**5. Changes/Problems:**

Nothing to report.

**6. Products:**

• **Publications, conference papers, and presentations**

○ **Journal Publications**

- Morgan, RK; Batra, H; Gaerig, VC; Hockings, J; **Brooks, TA**. "Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator", *BBA-Gene Regulatory Mechanisms*. 2015 Nov 18, 1859(2):235-245 e-pub ahead of print. PMID: 26597160

○ **Books or other non-periodical, on-time publications**

- Nothing to report

○ **Other publications, conference papers, and presentations**

- Morgan, R; Rahman, KM; **Brooks, TA**. Structure Elucidation of G-Quadruplex within the mid-region of the kRAS Promoter and Identification of Stabilizing Small Molecules as Promising Transcriptional Silencers. Proceedings of the American Association of Cancer Research (AACR), April 2015. Abstract 1245.
- Batra, H; **Brooks, TA**. The Effect of the transcription factor MAZ on kRAS transcription: a role for the G-quadruplex. Proceedings of the American Association of Cancer Research (AACR), April 2015. Abstract 2137.
- Morgan, R\*; Rahman, KM; **Brooks, TA**. Structure elucidation of G-quadruplex within the mid-region of the kRAS promoter and identification of stabilizing small molecules as promising transcriptional silencers. Mississippi Academy of Sciences Annual Meeting, Hattiesburg, MS. February 2015. \*1<sup>st</sup> place poster winner\*
- Batra, H; **Brooks, TA**. The Effect of the Transcription Factor MAZ on kRAS Expression: A Role in Pancreatic Cancer. ACS Drug Discovery and Development Colloquium, Little Rock, AR. 2014.
- Backus, K; **Brooks, TA**. A Novel Role for p53 in kRAS Binding in Pancreatic Cells. ACS Drug Discovery and Development Colloquium, Little Rock, AR. 2014

• **Website or other Internet site**

- Nothing to report

• **Technologies or techniques**

- Nothing to report

- **Inventions, patent applications, and/or licenses**

- Nothing to report

- **Other products**

- Research materials have been made through this funded project that includes a series of luciferase plasmids driven by various regions and mutations of the kRAS promoter. These are available to the scientific community by request, pending MTA arrangements, and are to be deposited with AddGene for dissemination.

## 7. Participants & Other Collaborating Organizations:

Name:	<i>Tracy Brooks</i>
Project Role:	<i>Primary Investigator</i>
Researcher Identifier (e.g. ORCID ID):	0000-0002-1100-2437
Nearest person month worked:	6
Contribution to Project:	<i>Dr. Brooks has overseen all aspects of the research projects and guided the GRAs in their studies.</i>
Funding Support:	<i>No change</i>

Name:	<i>Rhianna Morgan</i>
Project Role:	<i>Graduate Research Assistant</i>
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	12
Contribution to Project:	<i>Ms. Morgan is responsible for all structural characterization work, including all aspects of Aim 1.</i>
Funding Support:	<i>No change</i>

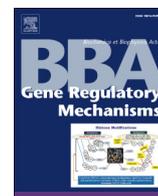
Name:	<i>Harshul Batra</i>
Project Role:	<i>Graduate Research Assistant</i>
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	12
Contribution to Project:	<i>Mr. Batra is responsible for all protein regulatory studies, including all aspects of Aim 2.</i>
Funding Support:	<i>No change</i>

## 8. Special Reporting Requirements:

There are no special requirements.

9. **Appendices:**

There are no appendices to attach.



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

Rhianna K. Morgan<sup>a</sup>, Harshul Batra<sup>a</sup>, Vanessa C. Gaerig<sup>b</sup>, Jennifer Hockings<sup>a</sup>, Tracy A. Brooks<sup>a,\*</sup>

<sup>a</sup> School of Pharmacy, Department of BioMolecular Sciences, Division of Pharmacology, University of Mississippi, University, MS 38677, USA

<sup>b</sup> College of Pharmacy, Department of Pharmacology and Toxicology, The University of Arizona, Tucson, AZ 85721, USA

## ARTICLE INFO

### Article history:

Received 19 December 2014

Received in revised form 13 November 2015

Accepted 17 November 2015

Available online 18 November 2015

### Keywords:

G-quadruplex

kRAS

Transcriptional control

Cancer therapeutics

## 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<sub>near</sub>). 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 G4s. G4<sub>near</sub> is predominantly a tri-stacked structure with a discontinuous guanine run incorporated; G4<sub>mid</sub> 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<sub>mid</sub>, is a stronger repressor of promoter activity than G4<sub>near</sub>. Using the extensive guanine-rich region of the kRAS core promoter, and particularly the G4<sub>mid</sub> 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.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

The *kRAS* gene, located on chromosome 12 at p12.1 [1], encodes for the p21<sup>RAS</sup> (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<sup>G12D</sup> 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.

\* Corresponding author at: 307 Faser Hall, University, MS 38677, USA.  
E-mail address: [tabrooks@olemiss.edu](mailto:tabrooks@olemiss.edu) (T.A. Brooks).

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 G4 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' G4 formations and function within the *kRAS* promoter is 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 G4 is formed from the mid-G4 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-G4 has 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, and dual luciferase assay kits were purchased from Promega (Madison, WI). <sup>32</sup>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 to 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 to 100 °C (spectra recorded each 7 °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<sub>M</sub>*. 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 °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 50 V; 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 [ $\gamma$ -<sup>32</sup>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

**Table 1**  
Oligonucleotide sequences.

Name	5'–3' sequence
PSA primer	TCGACTCTAAGCAATGCGT
G4 <sub>near</sub>	AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG
G4 <sub>near</sub> DMS	TTTTTTTAGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGGT TTTTTT
G4 <sub>near</sub> FRET	[6-FAM]-AGGGCGGTGTGGGAAGAGGGAAGAGGGGGA GG-[TAMRA]
G4 <sub>mid</sub>	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC
G4 <sub>mid</sub> DMS	[6-FAM]-TTTTTTTCGGGAGAAGGAGGGGGCCGGCCGG GGCCGGCCGGGGAGGAGCGGGGGCCGGCTTTTTTT
G4 <sub>mid</sub> FRET	[6-FAM]-CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGC GGGGAGGAGCGGGGGCCGGCC-[TAMRA]
G4 <sub>far</sub> mut A	AAGGGGTGCTGGGGCGGTCTAGGTGCGAGCCGGGCC CGTTGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC
mut B	CGGGGAGAAGGAGGTGCGGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC
mut C	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC
mut D	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC
mut E	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGTGGA GGAGCGGGGGCCGGCC
mut F	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGTGGCCGGCC
mut G	CGGGGAGAAGGAGGGGGCCGGCCGGCCGGCCGGGGGA GGAGCGGGGGCCGGCC

addition of 1  $\mu\text{g}$  of calf thymus DNA, the solutions were subjected to dimethyl sulfate (0.3% DMS in 1% ethanol) for up to 18 min. Each reaction was quenched in 0.3 M NaOAc and 0.2 M  $\beta$ -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, resuspended in a 95% formamide loading dye, and heated to 95  $^{\circ}\text{C}$  for 5 min before snap-cooling on ice. The cleaved products were separated on a 16% sequencing gel. Gels of  $G4_{\text{near}}$  with  $^{32}\text{P}$ -DNA were then dried and exposed to a phosphorimager and imaging was performed with a Storm 820 phosphorimager. Footprinting gels for FAM-labeled  $G4_{\text{mid}}$  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 Fig. 3.

### 2.6. FRET melt assay

Thermodynamic stabilities of  $G4$  DNA 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  $G4_{\text{near}}$  and Midland Certified Reagent Company, Midland, TX for  $G4_{\text{mid}}$ ). Stock concentrations of DNA were made in autoclaved, nuclease-free water at an approximate concentration of 100  $\mu\text{M}$ . 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  $^{\circ}\text{C}$  for 10 min, and allowed to anneal to room temperature slowly. Annealed DNA probe was then plated with 1  $\mu\text{M}$  compound in a 96 well plate; fluorescence was recorded every 1  $^{\circ}\text{C}$  from 25 to 95  $^{\circ}\text{C}$  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 1  $\times$  penicillin/streptomycin solution at 37  $^{\circ}\text{C}$ , in a humidified atmosphere containing 5%  $\text{CO}_2$ , in exponential growth. For cellular viability assays, cells were seeded in 96-well plates at a concentration of  $7.5 \times 10^3$  cells per well in 90  $\mu\text{L}$  of media, and allowed to attach overnight. The following day, a  $10 \times$  stock plate of TMPyP2 or TMPyP4 diluted from 5 mM over a 5–6 log range in 0.5 log steps was made and 10  $\mu\text{L}$  of this stock was added to the cell plate, in triplicate. The final high dose range was 0.08–500  $\mu\text{M}$ . Cell-free wells with the same dose-range were plated, and served as measurements of background absorption. 48 h later, 20  $\mu\text{L}$  of MTS + 5% PMS was added to each well and incubated for ~2 h 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  $\mu\text{L}$  of fresh media before the addition of MTS as above. Background absorptions were subtracted, and data were normalized to control cells;  $\text{IC}_{50}$  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  $2 \times 10^5$  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  $\mu\text{M}$  TMPyP2 or TMPyP4. 48 h 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  $\Delta\Delta\text{Cq}$  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 was constructed on the backbone of pGL4.17 plasmid (Promega). The promoter regions of interest, as denoted in Fig. 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).

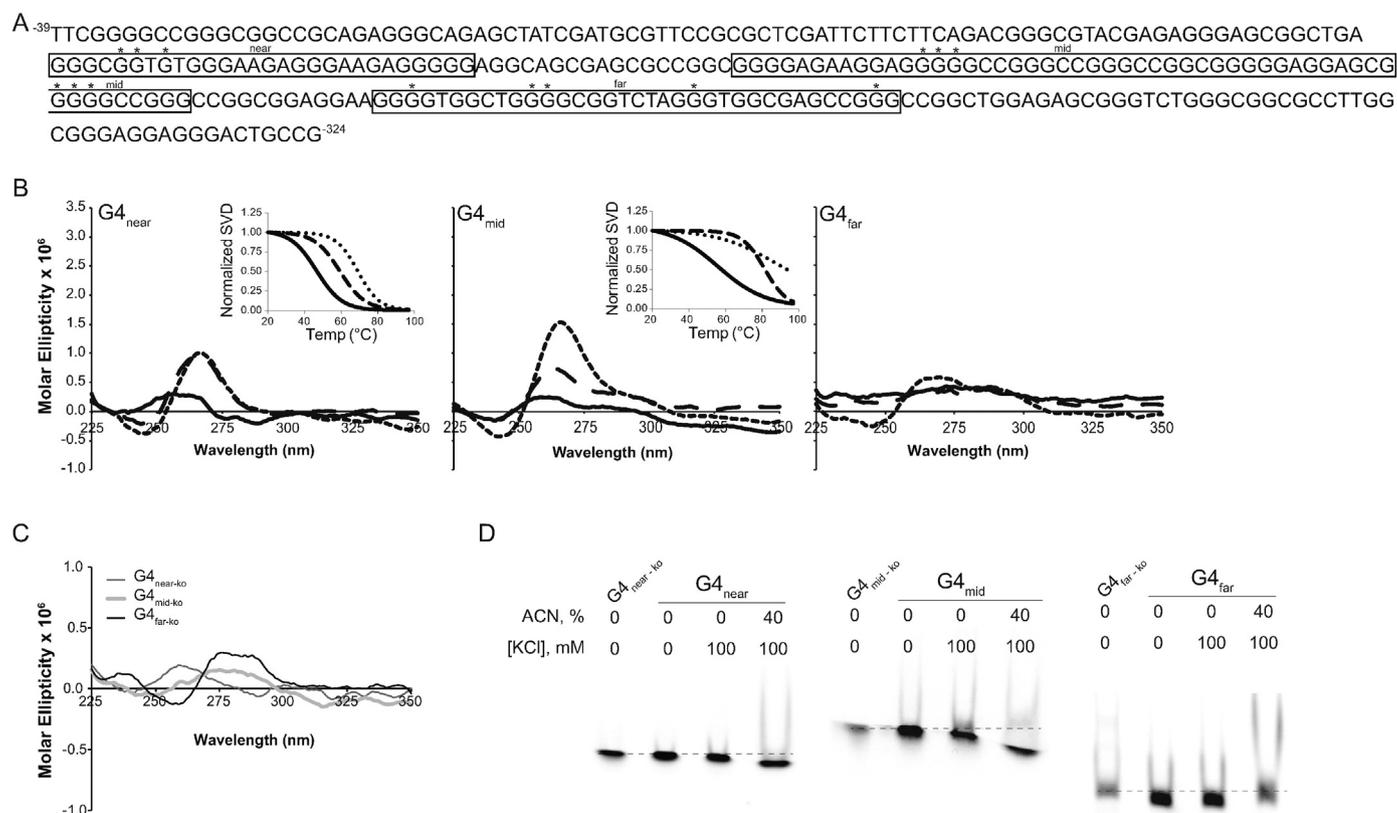
### 2.10. Transfection and luciferase assays

HEK-293 cells were maintained in Dulbecco's minimal essential medium supplemented with 10% fetal bovine serum and 1  $\times$  penicillin/streptomycin solution at 37  $^{\circ}\text{C}$ , in a humidified atmosphere containing 5%  $\text{CO}_2$ , in exponential growth. Before transfection with luciferase plasmids, cells were seeded in six-well plates at a concentration of  $2 \times 10^5$  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  $^{\circ}\text{C}$  in a 5%  $\text{CO}_2$  humidified environment overnight before media was replaced with DMEM  $\pm$  TMPyP2 or TMPyP4 in dark conditions, as the compound may be activated to a photosensitizer. After 48 h 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  $-510$  to  $+50$  bp surrounding the TSS [9–11], and contains cis-regulatory elements [34] and silencing  $G4$  elements [33]. Examination of the entire 500 bp upstream promoter revealed two previously undescribed putative  $G4$ -forming regions in addition to the previously described region [20,21,33] (Fig. 1A). Herein, we term the known 32-bp sequence, found  $-128$  bp from the TSS, as the near region, forming  $G4_{\text{near}}$ , the next distal region from  $-174$  to  $-226$  as the mid-region, forming  $G4_{\text{mid}}$ , and the furthest region from  $-238$  to  $-273$  as the far region, potentially forming  $G4_{\text{far}}$ . 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.



**Fig. 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 to 100 °C is 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-out (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<sub>near</sub> and G4<sub>mid</sub> 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<sub>far</sub> 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.

G4 formation and overall stability of the near-, mid-, and far-regions were examined by CD, each in the absence and presence of intramolecular G4-stabilizing KCl alone or with 40% anhydrous acetonitrile (Fig. 1B). As noted in previous publications, G4<sub>near</sub> 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<sub>mid</sub>) 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 that 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.

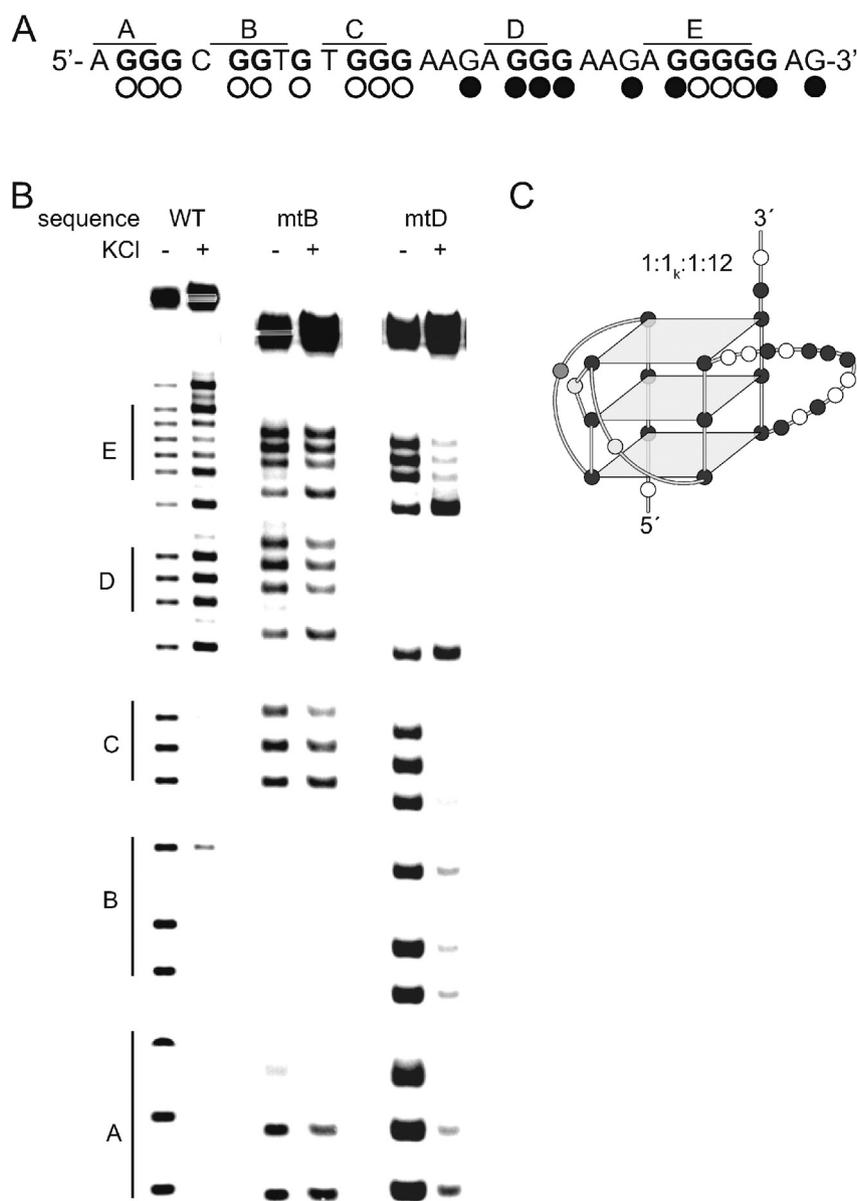
G-to-T knockout mutations were introduced into each G4-forming region (Table 1), and their higher order DNA formations were examined by CD in the presence of 100 mM KCl (Fig. 1C). As predicted the G-to-T mutations disrupted (G4<sub>near-ko</sub>, mutation confirmed by DMS footprinting below) or abrogated (G4<sub>mid-ko</sub> and G4<sub>far-ko</sub>) 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 (Fig. 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<sub>far</sub> 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<sub>far</sub> 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<sub>near</sub> 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 (Fig. 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 (Fig. 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



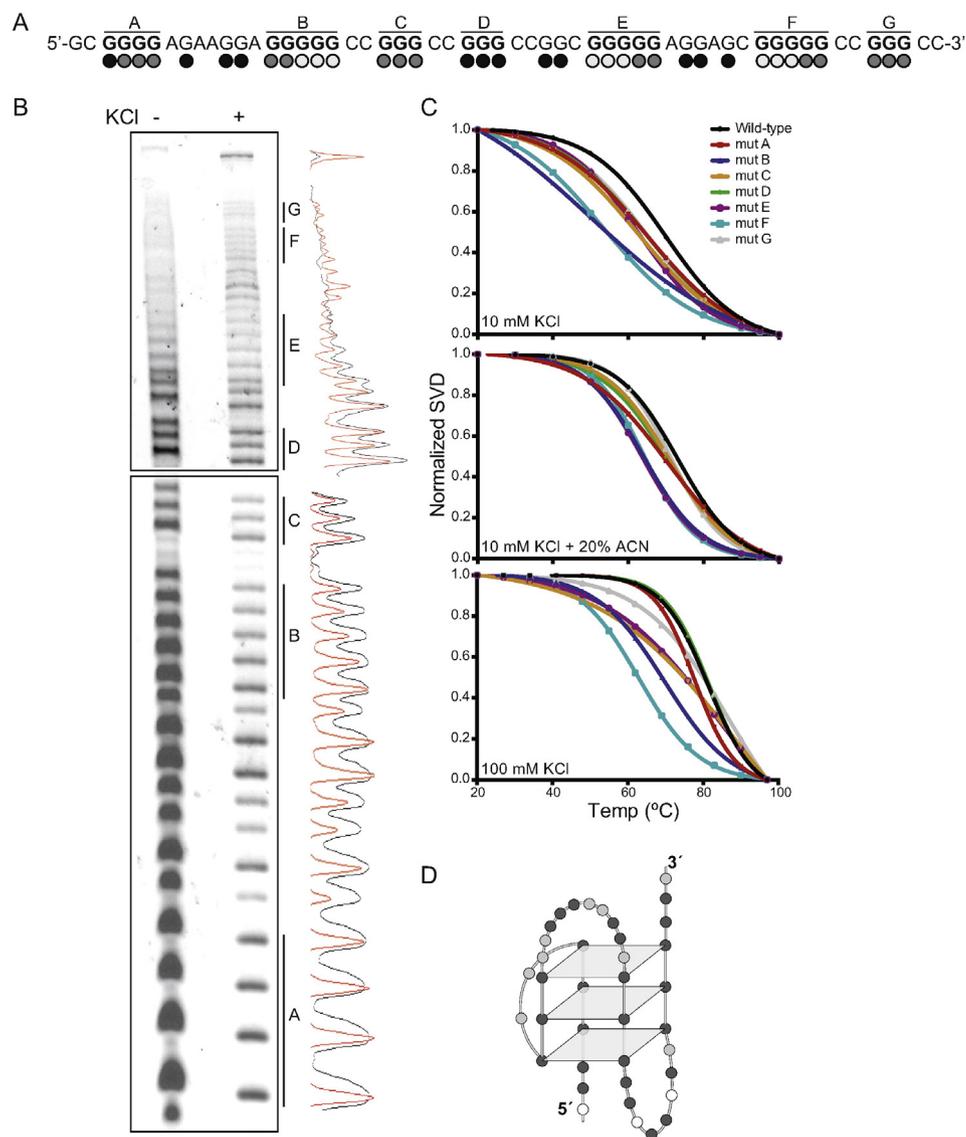
**Fig. 2.** Predominant G4 isoforms formed within the near kRAS promoter. (A) The  $G4_{\text{near}}$  sequence contains four runs of continuous guanines – A, C–E – and one discontinuous run – B. The open circles indicate evidence of protection in the subsequent DMS footprinting, whereas the black circles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

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 (Fig. 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 (Fig. 2B). These data, cumulatively with the CD findings from Fig. 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 (Fig. 2C). Collectively these

data support that the kinked structure shown in the literature [23] is indeed the predominant isoform within the entire  $G4_{\text{near}}$ -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 (Fig. 3A). DMS footprinting was performed on the entire wild-type  $G4_{\text{mid}}$  sequence without and with 100 mM KCl; histograms of the banding patterns were obtained with ImageJ software. It is clear that a number of higher order structures



**Fig. 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 circles 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, ImageJ 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 was 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.

are forming in equilibrium, with at most only partial protection patterns clear in runs B, E and F (Fig. 3B).

To aid in the preliminary description of the  $G4_{mid}$  structure, a series of single run knockout G-to-T mutations was 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 cations and dehydration stabilities in order to probe the most relevant guanine runs involved in the predominant structure (Fig. 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 (Fig. 3C). Melting profiles were examined in 100 mM KCl + 40% ACN conditions as well, but due to the extremely high thermal stability,  $T_M$ s 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

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

	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

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 (Fig. 3D).

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

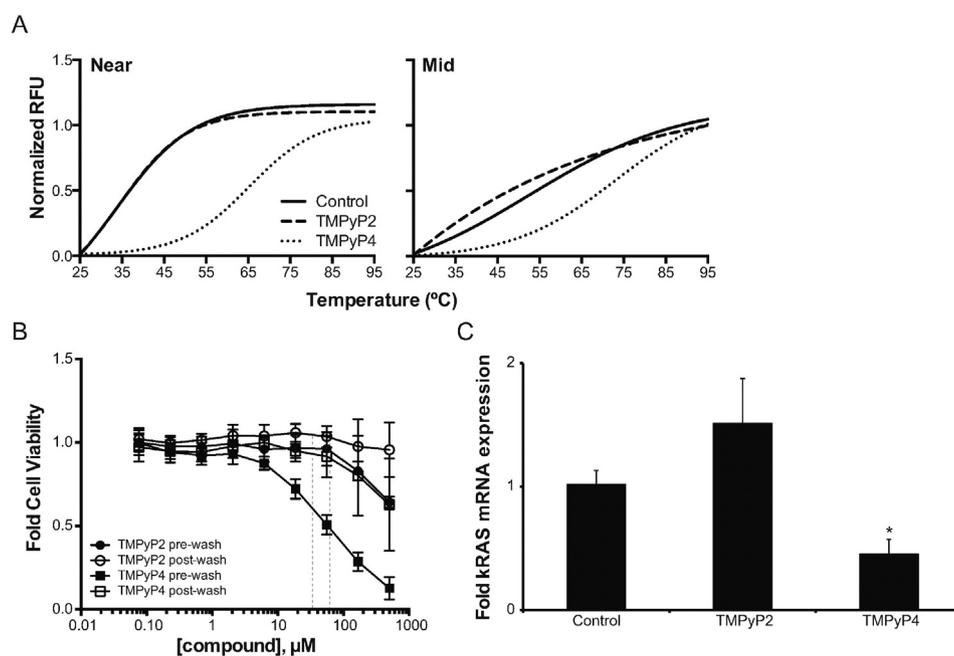
Previous literature has described kRAS promoter silencing in the presence of TMPyP4 [20,33], which we used along with its positional isomer TMPyP2 to evaluate the biological function of the near- and mid-G4-forming regions of the kRAS promoter. The effect of each cationic porphyrin on the thermal stability of the G<sub>4</sub><sub>near</sub> and G<sub>4</sub><sub>mid</sub> structures was determined by FRET melt (Fig. 4A). Consistently, 1  $\mu$ M TMPyP2 did not stabilize either structure, while 1  $\mu$ M TMPyP4 did. In particular, the T<sub>M</sub> of G<sub>4</sub><sub>near</sub> was 34 °C alone and in the presence of TMPyP2, but increased to 65 °C in the presence of TMPyP4. Similarly, G<sub>4</sub><sub>mid</sub>'s T<sub>M</sub> 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  $\mu$ M incubated with the cells for 48 h. 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  $\mu$ M,

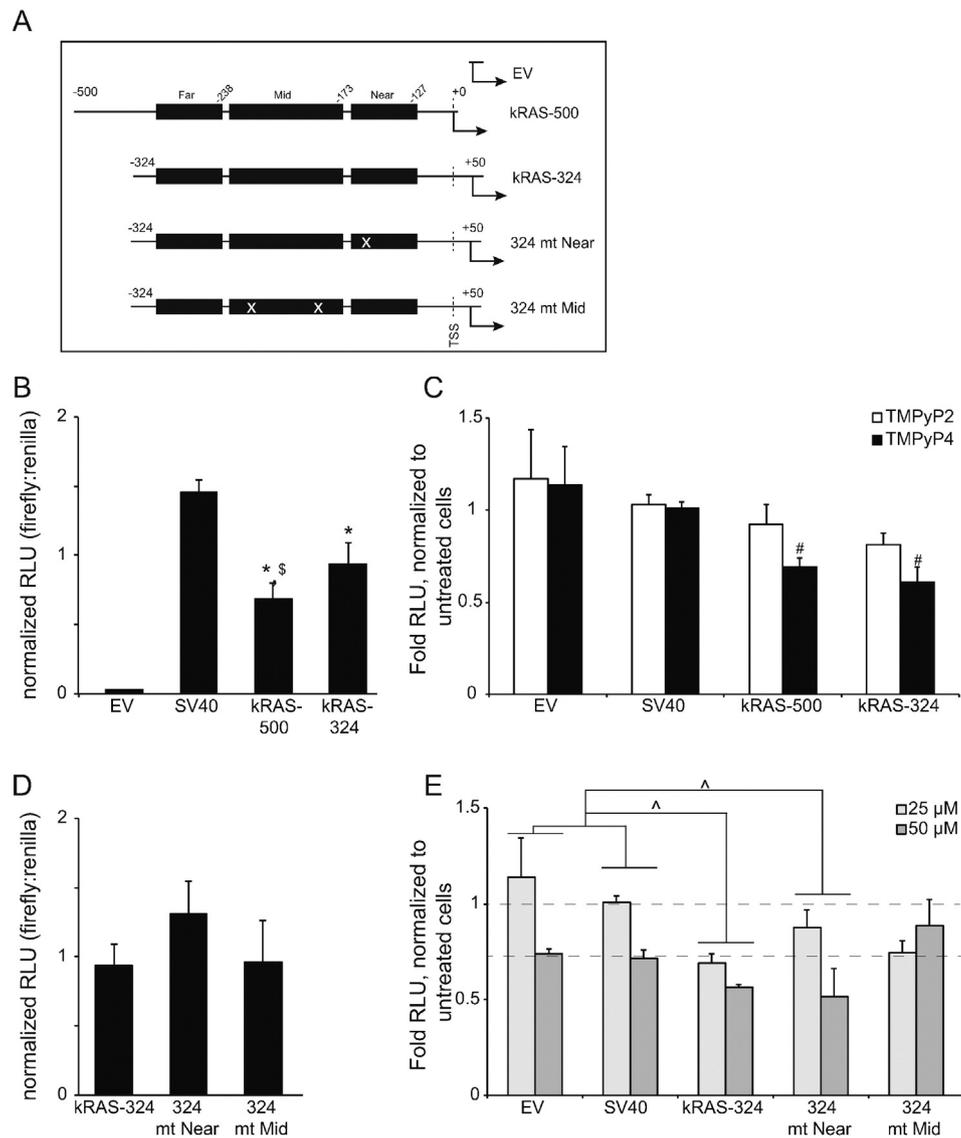
there is no effect on Panc-1 cellular viability. Moreover, at concentrations up to 500  $\mu$ M, TMPyP2 does not significantly impact Panc-1 cells, and TMPyP4 only decreases viability by approximately 25% (Fig. 4B). Panc-1 cells were incubated with 25  $\mu$ M TMPyP2 or TMPyP4 for 48 h, and changes in 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 (Fig. 4C). These findings support the role of G4-mediated silencing of kRAS expression.

In order to assess the contribution of G<sub>4</sub><sub>near</sub> and G<sub>4</sub><sub>mid</sub> to the silencing of kRAS transcription, a series of luciferase promoters was constructed (Fig. 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 to +0 in relation to the TSS, which had been shown to contain silencing G4s [33]. In addition a plasmid was assembled containing –324 to +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 Fig. 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 h, and firefly expression was normalized to pRL-driven renilla luciferase expression (Fig. 5B). SV40-promoter driven expression is almost 50-fold greater than activity from the promoterless EV ( $1.46 \pm 0.08$  versus  $0.03 \pm 0.01$ , 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  $0.69 \pm 0.11$  and kRAS-324 was  $0.94 \pm 0.16$ . Each of these plasmids was also transfected into HEK-293 cells, which were then exposed to 25  $\mu$ M TMPyP2 or TMPyP4



**Fig. 4.** Effect of G4 stabilization on pancreatic cancer cell viability and kRAS mRNA expression. (A) FRET Melt demonstrated a marked stabilization of both the G<sub>4</sub><sub>near</sub> and the G<sub>4</sub><sub>mid</sub> formations by the pan-G4-stabilizing TMPyP4 (1  $\mu$ M), but not the inactive positional isomer TMPyP2 (1  $\mu$ M). (B) The effect of each compound on pancreatic cancer cell, Panc-1, viability was measured at 48 h. 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  $\mu$ M, decreased cellular viability by 50%. Doses used in subsequent cellular assays are denoted by vertical dotted lines at 25 and 50  $\mu$ M. (C) Panc-1 cells were incubated with 25  $\mu$ M of either TMPyP2 or TMPyP4 for 48 h; subsequently kRAS mRNA, as normalized to GAPDH, was measured. TMPyP4, but not TMPyP2, significantly ( $*p < 0.05$ ) decreased transcription and supported 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).



**Fig. 5.** Isolating silencing G4 formations within the *KRAS* promoter. (A) A series of luciferase plasmids was constructed from the promoterless empty vector (EV) pGL4.17 backbone to localize the silencing G4s within the *KRAS* promoter. These included *KRAS*-500 (−500 to +0, relative to the transcriptional start site (TSS)), and *KRAS*-324 (−324 to +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 h 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 h) with 25  $\mu$ 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 h 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  $\mu$ 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  $\mu$ 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 ( $^{\wedge}p < 0.05$  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.

for 48 h (Fig. 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 was 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 Fig. 1C), while minimizing disruptions to potential transcription factor binding sites. Basal expression was measured from each of the 324 plasmids after 48 h of transfection, and there were no significant differences between any plasmids (Fig. 5D). In particular, as compared to the *KRAS*-324's expression of  $0.94 \pm 0.16$ , the expression for 324 mt Near and 324 mt Mid was  $1.31 \pm 0.23$  and  $0.96 \pm 0.30$ , respectively. Upon the addition of TMPyP2, as for all plasmids described above, no significant changes occurred with any plasmid with fold RLU of  $0.99 \pm 0.10$  and  $1.01 \pm 0.08$  for 324 mt Near and 324 mt Mid, respectively.

Upon exposure of the plasmids to 25 and 50  $\mu\text{M}$  TMPyP4, a number of significant effects were noted (Fig. 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  $\mu\text{M}$  compound to  $0.75 \pm 0.05$  and  $0.72 \pm 0.04$ , respectively. The dose–response noted in EV and SV40 plasmids was comparable, and was 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  $\mu\text{M}$  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  $0.55 \pm 0.09$  and  $0.51 \pm 0.15$  for the kRAS-324 and 324 mt Near plasmids, respectively, and was  $0.89 \pm 0.14$  for the 324 mt Mid plasmid. When the fold RLU 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  $0.56 \pm 0.13$ , for a 24% decrease as compared to the EV and SV40 plasmid non-specific effects (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  $G4_{\text{mid}}$  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 that 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  $\pm 50$  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  $\pm 50$  surrounding the TSS contains one consensus Sp1 binding site, five E2F-1 sites (four of which are in the critical region from 0 to +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, 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 Xodo's 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  $G4$ s 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 by [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  $G4_{\text{near}}$  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  $G4_{\text{near}}$  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  $G4_{\text{near}}$ , in addition to  $G4_{\text{mid}}$ , knockout mutations within the entire core kRAS promoter. While initial reports indicated a role of  $G4_{\text{near}}$  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 G4s, 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 have harbored this dysregulated oncogene.

Putative G4 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 compositions of G4s allow for specific targeting, similar to targeting a protein with a unique tertiary structure. The targeting of G4 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.

## Transparency document

The Transparency document associated with this article can be found in the online version.

## Acknowledgment

U.S. Department of Defense PRCRP grant (to Brooks) CA130229, startup funds from the University of Mississippi (to Brooks).

## References

- [1] A.A. Adjei, Blocking oncogenic Ras signaling for cancer therapy, *J. Natl. Cancer Inst.* 93 (2001) 1062–1074.
- [2] B.B. Friday, A.A. Adjei, K-ras as a target for cancer therapy, *Biochim. Biophys. 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, *Adv. Cancer Res.* 102 (2009) 1–17.
- [4] Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, RAS oncogenes: weaving a tumorigenic web, *Nat. Rev. 11* (2011) 761–774.
- [5] S.R. Hingorani, E.F. Petricoin, A. Maitra, V. Rajapakse, C. King, M.A. Jacobetz, S. Ross, T.P. Conrad, 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, *Semin. Cancer Biol.* 13 (2003) 267–273.
- [7] E. Wickstrom, Oligonucleotide treatment of ras-induced tumors in nude mice, *Mol. Biotechnol.* 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, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 5242–5247.
- [9] J. Jordano, M. Perucho, Chromatin structure of the promoter region of the human c-K-ras gene, *Nucleic Acids Res.* 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 Res.* 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 Res.* 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, *Nat. Commun.* 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, *Nat. Biotechnol.* 33 (2015) 877–881.
- [18] C.K. Kwok, S. Balasubramanian, Targeted detection of G-quadruplexes in cellular RNAs, *Angew. Chem.* 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, *Nat. Commun.* 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 Res.* 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 Res.* 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, *Chem. Commun. (Camb.)* 47 (2011) 4965–4967.
- [23] M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagawa, L.E. Xodo, Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription, *Nucleic Acids Res.* 37 (2009) 2841–2853.
- [24] R. Hänsel, F. Löhr, S. Foldynová-Trantirková, E. Bamberg, L. Trantirek, V. Dötsch, The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions, *Nucleic Acids Res.* 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, *Top. Curr. Chem.* 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. Opin. Struct. Biol.* 7 (1997) 126–134.
- [30] R.J. DeSa, I.B. Matheson, A practical approach to interpretation of singular value decomposition results, *Methods Enzymol.* 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, CA, U. S.) 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, *Environ. Health Perspect.* 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, *Sci. Rep.* 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 Res.* 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, *J. Med. Chem.* 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 Res.* 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, Switz.) 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, *Biochem. Biophys. Res. Commun.* 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, *J. Am. Chem. Soc.* 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, *J. Med. Chem.* 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, *J. Biol. Chem.* 285 (2010) 22003–22016.
- [43] J. Bendell, R.M. Goldberg, Targeted agents in the treatment of pancreatic cancer: history and lessons learned, *Curr. Opin. Oncol.* 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, *Investig. 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 Res.* 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, *Nat. Rev.* 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, *Investig. 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.