Author's response to reviews

Title: Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription

Authors:

John A Smestad (smestad.john@mayo.edu)
Louis J Maher III (maher@mayo.edu)

Version: 3
Date: 15 July 2015

Author's response to reviews: see over
Dear Editor,

Thank you for returning reviewer comments on the above-cited manuscript “Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription” by Smestad and Maher. We appreciate the opportunity to revise the submission in response to the thoughtful referee comments.

We have responded to all of the suggestions in the revised manuscript. Revisions are itemized in the following response listing and shaded in the revised manuscript itself.

We believe that the manuscript has been substantially strengthened by the helpful reviewer comments and trust that the work will now be found acceptable for publication.

Respectfully,

L. James Maher, III Ph.D.
Professor
Biochemistry & Molecular Biology
Dean, Mayo Graduate School

maher@mayo.edu
507-284-9041
Please find the response to referee comments and suggestions below. Revisions have been highlighted in cyan in the revised manuscript submission.

Reviewer: Yuliang Wu

Reviewer's report:
G-quadruplex is a stable four-stranded DNA or RNA structure formed in guanine-rich regions, and implicated in transcription, translation, replication, recombination, chromatin structure and telomere maintenance. Many helicases have been reported to function in G4 metabolism, such as WRN, BLM, FANCJ (BRIP1), PIF1, RHAU (DHX36), XPD, and DNA2. The correlation between transcription regulation and helicase, presence and absence, has been examined by many lab in many systems. In current report, based on two microarray datasets: BS (Proc Natl Acad Sci U S A. 2014 Jul 8; 111(27): 9905:9910) and WS (Stem Cell Reports. 2014 Apr 8; 2(4): 534:546), author re-analyzed the correlation between PQS position and gene expression, and found that PQS have position- and strand-dependent correlations with transcription, namely PQS located in sense strand 140-270 bp of downstream of TSS will increase transcription, PQS located in upstream of TSS (either strand) will decrease transcription.

We appreciate this review. Please find below our response to the following remarks:

1- The major concern is what's new from this study. It has been known G4 in promoter affect gene expression; it has been known BLM and WRN helicase may regulate gene expression by resolving G4 structure. The new information drawn from this report is about strand and distance of PQS with TSS. But the issue is whether this conclusion can be applied to other system, such as yeast, and other G4-resolving helicase.

We appreciate the concern raised by the referee and are thankful for this opportunity to clarify the several novel aspects of this work.

As the reviewer points out, other studies have previous associated PQS abundance with altered gene expression in BLM- and WRN-deficient fibroblast lines. The reviewer is correct that our new study focuses specifically on PQS abundance in BS and WS cell lines, and strand localization of PQS and distance of PQS to the TSS in particular. Strand localization and distance to the TSS are crucially important, however, since we now clearly demonstrate that PQS abundance in BS and WS differentially-expressed genes varies strongly with position and strand. This is a novel result, and these previously-unappreciated patterns are intriguing in their own right as they further define the characteristics of genes transcriptionally sensitive to presence of BLM and WRN helicases.

In addition to our analysis of PQS abundance in BS and WS cell lines, we wish to point out that the present work also demonstrates for the first time that the PQS motifs in BS and WS differentially expressed genes are significantly biased in composition. This is extremely important because it suggests that certain types of PQS motifs are preferentially involved in regulation of gene expression through interaction with RecQ helicases. A motif classification analysis of PQS such as we present for genes transcriptionally sensitive to RecQ helicase presence has never before been reported. This approach may prove very useful in understanding the genome-regulatory biology of PQS motifs.

A third novel feature of the present work is our demonstration, for the first time, that PQS in multiple human cell lines are associated with activated or repressed transcription in a strand- and position-dependent manner. The only other work previously published in this area (Zhao et al. (2008) Genome Research), using what are now outdated techniques, identified that the region 0-500 bp downstream from the TSS is associated with increased transcription. This previously published study, due to its necessarily insensitive method, was unable to identify regions where PQS were correlated with decreased transcription. The data that we present are more sensitive and granular, using a 200 bp bin size, and covering a full 2 kbp upstream and
downstream of the TSS. Thus, our approach greatly improves the resolution of analysis for existing data sets. We therefore can propose a more granular model in which PQS on both strands upstream and downstream of the TSS are generally associated with repressed transcription, but in which PQS located between 140-270 bp downstream of the TSS on the sense strand are associated with increased transcription. This model in which PQS have at least two distinctly different transcriptional effects has never been proposed before. We believe this is the major finding of this present work.

In regard to other model systems, it is indeed fascinating to speculate that the findings of our study regarding G4-resolving helicases may hold true in other eukaryotic organisms such as yeast. Defects in the Sgs1 RecQ helicase in yeast successfully recapitulate many of the phenotypic and genomic characteristics of WRN and BLM-deficient cells including telomere defects and premature aging. Sgs1 also interacts with many proteins that have known orthologues in humans, and many of these orthologues have been verified as interaction partners for BLM and WRN. It is indeed possible that the close biological similarity between Sgs1, BLM, and WRN extends to the PQS pattern associations identified in this present work. We believe that analysis of PQS patterns in yeast is beyond the scope of the present work.

To help clarify this novelty for the reader we have added the following section to the concluding paragraph of the manuscript:

Particularly novel insights this these high-resolution analyses are that i) PQS abundance in BS and WS differentially-expressed genes varies strongly with both position and strand; ii) PQS motifs in BS and WS differentially expressed genes are significantly biased in composition; and iii) PQS in multiple human cell lines are associated with activated or repressed transcription in a strand- and position-dependent manner.

2- What’s the consequence of PQS located in antisense strand, downstream of TSS?

We thank the reviewer for this question and wish to direct his attention to Figure 2 D. Here we plot the prediction error of our epigenomic-based gene expression estimates (actual value minus prediction) as a function of PQS strand and position relative to the TSS. Values less than zero indicate that PQS at given position correlate with reduced transcription. For the antisense strand, all positions identified as statistically significant show reduced transcription. In specific regard to the consequences of PQS located in the antisense strand downstream of the TSS, our data show significant correlations of PQS with reduced expression between 0 and 760 bp downstream. Beyond 760 bp, any correlations that we identified were not statistically significant.

3- Another microarray dataset for WS and BS is also reported in Nucleic Acids Res. 2010 Mar; 38(4): 1114-1122. While author have not chosen this one, thus the choice of dataset is biased?

We appreciate the referee’s reference to this important literature and are thankful for the chance to respond to this interesting concern. The mentioned dataset published by Johnson et al. (2010) in Nucleic Acids Research reflects fibroblasts lines from three BS patients and three controls. The Bloom Syndrome differential gene expression dataset used in the current work and generated by Nguyen et al. (2010) PNAS was created from comparison of microarray gene expression profiles from fibroblast lines of 15 BS patients and 11 controls. In fact, this 15-patient sample dataset includes cells from the same three patients included in the Johnson et al. dataset, but is superior due to the statistical power from having more patient and control samples.

4- No any connection of epigenetic and gene expression data with BS and WS/G4 (Figure 2).

We appreciate this perceptive comment. The intention of Figure 2 is to communicate that PQS measurably affect gene expression in normal cells (not BS or WS) in a way that correlates with
strand and distance from TSS. Through understanding the transcriptional correlations of PQS strand and position, the goal is to be able to interpret the PQS abundance patterns observed in BS and WS cell lines, thereby combining two distinctly different types of data to provide a basis for implying biological function.

5- In microarray analysis, the cutoff of fold expression difference of #1.5 is pretty minor, although in PNAS BS paper they used 1.5, but for WS case (Stem cell reports) they used 2.0.

We appreciate this thoughtful comment. The referee is correct that two recently published papers use different fold expression threshold values. This highlights the fact that currently there is no statistically-valid argument for an appropriate fold expression change criterion that is considered significant. Low thresholds favor sensitivity at the cost of increasing the FDR, while higher threshold values lower FDR at the cost of sensitivity. We believe that our choice of 1.5 for fold-expression change was appropriate based on the prerogative of many other papers in this field (e.g. Kaur et al. (2012) J Clin Endocrinol Metab, Nguyen et al. (2014) PNAS, Kuo et al. (2013) J Nutrition, Cagnone et al. (2012) Biol Reproduct). This criterion, though common in the literature, is admittedly arbitrary.

6- Page 6 line 1, up-regulated gene number is 1012.

We thank the referee for noting this mistake. We have corrected this number on page 6, as well as on page 17, in Table S1 and in Table S3.

7- Figure 6 is very complicated, whether can be simplified?

We agree that the figure is very complicated and appreciate the opportunuity to simplify it in the resubmission. We have minimized text in the figure, added letter cues to the legend, and simplified the diagrams.

Specific comments for the manuscript:

- **Level of interest**: An article whose findings are important to those with closely related research interests
- **Quality of written English**: Acceptable
- **Statistical review**: Yes, but I do not feel adequately qualified to assess the statistics.
- **Declaration of competing interests**: 'I declare that I have no competing interests'

Reviewer: Joshua Sommers

Reviewer's report:
Major Compulsory Revisions: None

We appreciate this supportive review.

Minor Essential Revisions:

1.- Page 2 Line 16: since this is the first occurrence of TSS it should not be abbreviated

We thank the referee for pointing this out and have made the following correction to the text of the manuscript on page 2.

“To interpret these correlations we determined genome-wide PQS correlations with transcription while controlling for epigenetic status. Our results identify multiple discrete transcription start site-proximal positions where PQS are correlated with either increased or decreased transcription.”

2.- Page 4 Lines 8-12: BLM and WRN can unwind other structures as well, and these helicases
may have other important functions other than unwinding G4 structures, perhaps you could change the language to indicate that it is one of the helicases functions.

We appreciate the opportunity to clarify this aspect of BLM and WRN helicase function. Below is the text of the revised paragraph:

“The BLM and WRN helicases implicated in BS and WS, respectively, have both been shown to unfold G-quadruplex structures assembled in vitro [7, 8], in addition to other types of DNA structures. It has been proposed that BLM and WRN helicases function in vivo by resolving DNA structures, including intrastrand and interstrand G-quadruplexes hypothesized to form at putative G-quadruplex-forming sequences (PQS) during homologous recombination [9], base-excision repair (WRN only) [10], in telomeres during cellular replication [11-13], and in regulation of gene transcription [14].”

3.- Page 5 Line 3: can abbreviate transcription start sites to TSS here.

We thank the referee for raising this concern. However, we prefer not to use the abbreviation, “TSS”, in this case since this is the first usage of the phrase, “transcription start sites”, in the body of the manuscript.

4.- Page 7 Lines 16-18: please clarify, more than 500 bp upstream, downstream or both.

We appreciate the need for clarification and have modified the language in this section in the manuscript. Below is the text from the revised sentence.

“In that work, PQS from 0 to 500 bp downstream of TSS were correlated with increased gene expression. PQS from 0 to 500 bp upstream of the TSS were not significantly correlated with gene expression.”

5.- Page 12 Line 20-21: should be corrected to “one PQS within 2 kbp of a TSS”?

The reviewer is absolutely correct in this and we thank him for detecting this error. Below is the revised sentence in the manuscript.

“Interestingly, up- and down-regulated genes in BS and WS are more likely to have at least one PQS within 2 kbp of a TSS (BS: 84% of up-regulated genes and 90% of down-regulated genes, WS: 74% of up-regulated genes and 84% of down-regulated genes, genome-wide average: 55%; Additional file 1: Table S1)”

6.- Page 19 Line 19: are the parentheses required after “prop.test”?

We thank the referee for pointing this out and have removed the parentheses from this sentence to avoid confusing readers. The amended sentence now reads as follows:

“Test datasets and randomly-generated datasets were compared to the genome-wide dataset using a p-value generated from the prop.test function in R.”

7.- Figure 3: it would be clearer for readers if the legend had “PQS number enrichment” instead of just “number enrichment”, same for depletion.

We appreciate the referee for pointing out this opportunity to clarify. We have changed the figure legend to read “PQS number enrichment” and “PQS number depletion”.

8.- Figure S1: please explain how 2 genes are both up-regulated and down-regulated in WS cells.

We thank the referee for discovering this error and have updated Additional File 1 with a corrected version of the figure.
9.- Figure S2: y-axis labels should read “PQS frequency” for clarity.

We thank the referee for pointing this out and have made the change to the figure y-axis.

10.- Table S3: please correct 6.1E+06 to 6.1E-06.

We appreciate the referee for pointing out this error, which we have corrected in the revised table.

Discretionary Revisions:
1.- Table S2: would be nice to have this data in the manuscript and not just as a supporting file.

We appreciate the referee for this suggestion. After discussion and consideration we respectfully believe that, given its size and format, the best disposition of this table is as a supporting file.

2.- Table S4: please comment on cell line GM12878 analysis in the manuscript, seems like it has a quite high false discovery rate.

We completely agree that this point deserves clarification. The reason for the high FDR in this one cell line is unclear, but this value reflects that fewer PQS positions were identified to be significantly correlated with altered transcription (7 positions in sense strand, compared to the average of 59 positions in other cell lines; 8 positions in antisense strand, compared to average of 42 positions in other cell lines). The discussion section has been modified to contain the following clarifying paragraph:

“This analysis was repeated for each of 7 human cell lines, generating a composite analysis of all statistically-significant PQS positions correlated with gene expression (Figure 2D). False discovery rates (FDR) for the analysis of sense and antisense strands were below 10% for four of the seven cell lines tested, indicating that the observed correlations have very low probability of being statistical noise (Additional file 1: Tables S4). The GM12878 cell line notably had a very high FDR for both DNA strands (47% for antisense strand and 54% for sense strand). The reason for this is unclear, but the high FDR reflects the presence of fewer identified positions where PQS significantly correlated with altered transcription. This is the first PQS correlation analysis incorporating epigenetic data from multiple cell lines and extending the study region to 2 kbp upstream and downstream of TSS. Interestingly, the general agreement for data from different cell lines suggests that PQS position correlates with gene expression regardless of epigenetic context. The analysis correlates PQS position in the antisense strand with lower gene expression, regardless of PQS position upstream or downstream of the TSS. In contrast, PQS in the sense strand show different correlations depending on position. PQS in the sense strand are correlated with lower gene expression, except for PQS positioned downstream of the TSS between 140-270 bp, 1750-1770 bp, and 1900 bp. PQS at these three sense strand positions were correlated with increased gene expression. This analysis did not find PQS correlations with gene expression at all locations. In total, 33% and 35% of analyzed PQS positions in the antisense and sense strands, respectively, displayed statistically-significant PQS abundance correlated with gene expression. These data represent a large improvement over previous studies in terms of resolution and coverage.”

3.- As part of your discussion section, it would be very helpful to mention other G4 unwinding helicases such as FANCJ, as well as RECQ1 which binds to G4 DNA but does not seem to unwind them. How would an enzyme like RECQ1 be expected to affect transcription based on your model?

We appreciate this suggestion and are in agreement that this is a potentially fascinating topic. In response, we have added the following paragraph to the discussion section of the manuscript.
“It is intriguing to consider whether correlations of PQS abundance/location with transcriptional sensitivity to RecQ helicase loss observed in this study may similarly be found for other DNA helicases such as the RecQ4 helicase (deficient in Rothmund-Thomson photosensitivity- and cancer-associated syndrome), FANCJ helicase (deficient in Fanconi Anemia), and RecQ1 and RecQ5 helicases (members of the RecQ helicase family, but lacking evidence of a human phenotype). Of these helicases, BLM, WRN, RecQ4, and FANCJ reportedly have measurable affinity and helicase activity for G-quadruplex DNA in vitro, in addition to other diverse specificities and activities that are unique to each helicase. It is possible that similar patterns to those observed for BLM and WRN helicases in the present work may also be observed for RecQ4 and FANCJ helicases, although testing this experimentally is beyond the scope of the present work. There is less experimental evidence, however, that RecQ1 and RecQ5 have significant affinity and helicase activity for G-quadruplex DNA, although these helicases do have affinity and helicase activity for other DNA structures. It would therefore be less likely that significant correlation would be observed between PQS abundance/location and transcriptional change in RecQ1 and RecQ5-deficient cell lines.”

Level of interest: An article of importance in its field  
Quality of written English: Acceptable  
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.  
Declaration of competing interests: I declare that I have no competing interests