

QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences

Oleg Kikin, Lawrence D'Antonio and Paramjeet S Bagga*

Bioinformatics, School of Theoretical and Applied Science, Ramapo College of New Jersey, Mahwah, NJ 07430, USA

Received February 15, 2006; Revised March 1, 2006; Accepted March 30, 2006

ABSTRACT

The quadruplex structures formed by guanine-rich nucleic acid sequences have received significant attention recently because of growing evidence for their role in important biological processes and as therapeutic targets. G-quadruplex DNA has been suggested to regulate DNA replication and may control cellular proliferation. Sequences capable of forming G-quadruplexes in the RNA have been shown to play significant roles in regulation of polyadenylation and splicing events in mammalian transcripts. Whether quadruplex structure directly plays a role in regulating RNA processing requires investigation. Computational approaches to study G-quadruplexes allow detailed analysis of mammalian genomes. There are no known easily accessible user-friendly tools that can compute G-quadruplexes in the nucleotide sequences. We have developed a web-based server, QGRS Mapper, that predicts quadruplex forming G-rich sequences (QGRS) in nucleotide sequences. It is a user-friendly application that provides many options for defining and studying G-quadruplexes. It performs analysis of the user provided genomic sequences, e.g. promoter and telomeric regions, as well as RNA sequences. It is also useful for predicting G-quadruplex structures in oligonucleotides. The program provides options to search and retrieve desired gene/nucleotide sequence entries from NCBI databases for mapping G-quadruplexes in the context of RNA processing sites. This feature is very useful for investigating the functional relevance of G-quadruplex structure, in particular its role in regulating the gene expression by alternative processing. In addition to providing data on composition and locations of QGRS relative to the processing sites in the pre-mRNA sequence, QGRS Mapper features interactive graphic

representation of the data. The user can also use the graphics module to visualize QGRS distribution patterns among all the alternative RNA products of a gene simultaneously on a single screen. QGRS Mapper can be accessed at <http://bioinformatics.ramapo.edu/QGRS/>.

INTRODUCTION

The quadruplex structures formed by guanine-rich nucleic acid sequences have received significant attention recently because of increasing evidence for their role in important biological processes and as therapeutic targets (1–5). The G-quadruplex structure, also known as a G-quartet, is formed by repeated folding of either the single polynucleotide molecule or by association of two or four molecules. The structure consists of stacked G-tetrads, which are square co-planar arrays of four guanine bases each (6). G-quadruplex is stabilized with cyclic Hoogsteen hydrogen bonding between the four guanines within each tetrad. The present work focuses only on the unimolecular quadruplexes, since it is more likely to be encountered in physiological conditions (7,8).

Guanine-rich sequences capable of forming G-quadruplexes are found in telomeres, promoter regions, transcribed and other biologically important regions of the mammalian genomes. G-quadruplex DNA has been suggested to regulate DNA replication in retinoblastoma susceptibility gene (Rb) region (9). This structure may control cellular proliferation at telomeric level and by transcriptional regulation of oncogenes like *c-myc* (2,10,11) and *c-kit* (12). Formation of G-quadruplex seems to be regulated through interactions with cellular proteins. While some proteins help stabilize the structure (13), others are known to resolve it (1,4,14,15). Proteins and chemicals that stabilize the G-quadruplex structure can inhibit telomerase action and, therefore, are being evaluated as anticancer therapeutic agents (16–20). Chemical compounds that inhibit G-quadruplex helicase activity may also be capable of regulating cellular proliferation (4). G-quadruplexes are also being eyed as potential antimicrobial agents due to their ability to transport monovalent anions (21).

*To whom correspondence should be addressed. Tel: +1 201 684 7722; Fax: +1 201 684 7637; Email: pbagga@ramapo.edu

G-quadruplex motifs in the RNA have been shown to play significant roles in mRNA turnover (1), FMRP binding (22), translation initiation (23) as well as repression (24). We have shown previously that a G-rich sequence (GRS) can mediate 3' end processing of mammalian pre-mRNAs by interacting with DSEF-1/hnRNP/H' protein (25–27). Members of the hnRNP H protein family recognize G-rich motifs capable of forming G-quadruplexes and are known to regulate polyadenylation and splicing events in mammalian transcripts (28–30). Regulated RNA processing is an essential component of differential gene expression which is central to many important biological processes. More than half of human genes are known to have alternative polyadenylation (31). Over two-thirds of human genes are thought to undergo alternative splicing (32). Sequences capable of forming G-quadruplexes found in the vicinity of polyadenylation and splice sites act as regulators by interacting with hnRNP F and H proteins (25–27,30,33). Whether quadruplex structure directly plays a role in regulating RNA processing events requires investigation.

Computational approaches to study G-quadruplexes in the mammalian genomes allow large-scale and detailed analysis of mammalian genes. Although, G-quadruplexes have been surveyed in the human genome with such techniques (34,35), there are no known user-friendly computational tools easily accessible to the public. We had previously built a database of mapped G-quadruplex sequences in selected alternatively processed human and mouse genes (36). Our preliminary analysis of the database suggests prevalence of such motifs near alternative splice and poly(A) sites. We have now developed a web-based server, QGRS Mapper, that generates detailed information on composition and distribution of putative quadruplex forming G-rich sequences (QGRS) in any NCBI nucleotide sequence identified or provided by the user. The program is also designed to handle the analysis of mammalian pre-mRNA sequences, including those that are alternatively processed (alternatively spliced or alternatively polyadenylated). Researchers interested in predicting the ability of a nucleotide sequence to form G-quadruplex structure will find QGRS Mapper to be a user-friendly application that provides many options for analysis. The user can define the minimum number of tetrads, maximum length of the G-quadruplex motif, and size as well as composition of the loops. The program can map unimolecular QGRS in the entire nucleotide sequence provided in the raw or FASTA format by the user. This method can be used for the analysis of genomic sequences, e.g. promoter and telomeric regions, as well as RNA sequences. It is also useful for predicting G-quadruplex structures in oligonucleotides. Alternatively, the program provides options to search the entire NCBI Gene Entrez, RefSeq and GenBank databases in order to retrieve the desired gene/nucleotide sequence entries for analysis of their transcribed regions. Furthermore, QGRS Mapper is a unique tool for mapping G-quadruplex forming sequences in the context of RNA processing sites. This feature is very useful for investigating the functional relevance of G-quadruplex structure, in particular its role in regulating the gene expression by alternative processing. In addition to providing data on composition and locations of QGRS relative to the processing sites in the pre-mRNA sequence, QGRS Mapper offers interactive graphic representation of the data.

The user can also use a graphics module to visualize QGRS distribution patterns among all the alternative RNA products of a gene simultaneously on a single screen.

METHODS

QGRS definition

The main goal of the QGRS Mapper program is to predict the presence of QGRS in nucleotide entries. These putative G-quadruplexes are identified using the following motif.



Here x = number of guanine tetrads in the G-quadruplex and y_1, y_2, y_3 = length of gaps (i.e. the length of the loops connecting the guanine tetrads). The motif consists of four equal length sets of guanines (which we call G-groups), separated by arbitrary nucleotide sequences, with the following restrictions.

- The sequence must contain at least two tetrads (i.e. $x \geq 2$). Although structures with three or more G-tetrads are considered to be more stable, many nucleotide sequences are known to form quadruplexes with two G-tetrads (37,38). QGRS Mapper is meant to be a flexible and comprehensive tool for investigating G-quadruplexes; hence it considers sequences with two tetrads.
- By default, only QGRS of maximum length of 30 bases are considered. However, the program gives the user the option to search for sequences up to 45 bases. This restriction on the length of the sequences being considered is in agreement with recent literature (34,35). The maximum length of 30 bases restricts G-groups to a maximum size of 6.
- The gaps or loops between the G-groups may be arbitrary in composition or length (within the overall restrictions on the length of QGRS). The program gives the user the option to search for QGRS having loops with a specified length range (e.g. the user can search for QGRS with loops of lengths between 1 and 4). The user can also specify a string that one or more loops of each QGRS must contain. This string can be given as a regular expression. For example, entering the regular expression 'T{3,5}' will search for QGRS having one or more loops that contain three to five consecutive T's.
- Also, at most one of the gaps is allowed to be of zero length

Table 1 shows some examples of valid QGRS. The guanine groups which form the tetrads are underlined.

The first sequence has four tetrads and equal length gaps. This would seem to provide a G-quadruplex that is the most stable of the three sequences. The second sequence is notable for the significant differences in the size of its loops. The third sequence has two tetrads, even though three of

Table 1. QGRS examples (G-groups are underlined)

| Sequence | QGRS parameters |
|----------------------------------|-------------------------------------|
| <u>GGGGTGGGGTGGGGTGGGG</u> | $x = 4, y_1 = 1, y_2 = 1, y_3 = 1$ |
| <u>GGGTGGGTGGCAGAGCTGGGCTGGG</u> | $x = 3, y_1 = 1, y_2 = 10, y_3 = 2$ |
| <u>GGGCGGGCTGGGTTGG</u> | $x = 2, y_1 = 2, y_2 = 3, y_3 = 3$ |

the G-groups could have included another G (since all G-groups must be equal in size).

G-scores

We have devised a scoring system that evaluates a QGRS for its likelihood to form a stable G-quadruplex. Higher scoring sequences will make better candidates for G-quadruplexes. The scoring method uses the following principles which are based on previous studies (34,35,39–42).

- Shorter loops are more common than longer loops.
- G-quadruplexes tend to have loops roughly equal in size.
- The greater the number of guanine tetrads, the more stable the quadruplex.

The computed G-scores are dependent on the user selected maximum QGRS length. The highest possible G-score, using the default maximum QGRS length of 30, is 105. Here is a sequence attaining that score:

GGGGGGTGGGGGGTGGGGGGTGGGGGG.

Eliminating QGRS overlaps

Two QGRS are said to overlap if their positions in the nucleotide sequence overlap. QGRS Mapper will start with a nucleotide sequence, find all QGRS occurring in the sequence and then produce a non-overlapping set of QGRS. Overlaps are eliminated by selecting the higher scoring QGRS. In the non-overlapping view, only this sequence will be displayed, although the user can request that all overlapping sequences be displayed.

Please see supplementary materials at NAR online for more details on the elimination of overlapping QGRS.

FEATURES

Design and implementation

QGRS Mapper is a web-based program, written in PHP, with Java being used for some of its graphics. The program takes a nucleotide sequence from NCBI (or as provided by the user) and analyzes it for the presence of putative G-quadruplexes. The structure of QGRS Mapper is summarized in Table 2.

Search and analysis

QGRS Mapper allows the user to search for putative G-quadruplexes in a variety of ways. It is possible to enter

a nucleotide sequence in raw or FASTA format for analysis. One can search and analyze gene sequences by Gene ID, Gene name or symbol, accession number or GI number for an NCBI nucleotide sequence entry. The user can opt to change the maximum length of QGRS that will be searched for (the default maximum length being 30) and change the minimum sized G-group (which is two by default). Also, the user can specify that the loops in the QGRS fall within a given size range and that one or more loops of the QGRS contain a given string (for which the user may enter a regular expression). The web page for QGRS analysis can be seen in Figure 1.

After entering a sequence in raw or FASTA format, QGRS Mapper will search the sequence for occurrences of QGRS. The user may enter any combination of the letters A, C, T, G, U, N.

The Gene ID field allows the user to search the NCBI Entrez Gene database. QGRS Mapper will connect to NCBI, download and parse the gene entry, and then analyze the transcribed region of its nucleotide sequence for the presence of QGRS. For example, entering the gene ID 403437 results in downloading the Brca1 gene sequence for *Canis familiaris*. Using the default QGRS search parameters, QGRS Mapper finds 156 non-overlapping QGRS and 3394 overlapping QGRS in the transcribed region of this gene.

The Gene Name or Gene Symbol field also allows the user to search the NCBI databases for all such genes. Entering the gene name Bcl2 results in nine different hits which are displayed in Table 3.

All nine of these entries can be analyzed for the occurrence of QGRS. Clicking on the Gene ID takes the user to the respective Entrez Gene entry. Clicking on the last column initiates analysis of the selection by QGRS Mapper.

Similarly, the user can also enter an NCBI accession number to search for gene sequences. For example, searching the accession number AF312033 results in 12 hits being displayed for this GenBank nucleotide sequence entry.

The search phase of the program is followed by an analysis of the QGRS contained in the query sequence. In this phase of QGRS Mapper, the sequence data downloaded previously is analyzed to identify and map all QGRS relative to locations such as splice sites in exons/introns, and poly(A) site (if these locations are known). Furthermore the QGRS are scored by the method described above. The computed G-score is used to eliminate overlapping QGRS.

At times, QGRS Mapper must analyze a considerable amount of data. For example, the mouse version of the gene PTPRU, which is 69822 bases long, contains 94681 QGRS of length up to 45 bases. QGRS Mapper will find, analyze and map all of these sequences. During this analysis a message is displayed indicating the estimated time left to completion.

QGRS Mapper output

After the analysis of overlaps is completed, QGRS Mapper displays a summary of its findings, in the Gene View. This summary includes basic gene information such as the gene ID, gene symbol, gene name, a link to the NCBI entry, organism name, chromosome number and number of products and poly(A) signals. Information is also given for each product, such as the number of exons and introns, number of QGRS

Table 2. QGRS Mapper

| Component | Function |
|-----------------|--|
| Retrieve | Query, retrieve and parse gene information from NCBI or obtain nucleotide sequence from user. |
| Search | Identify all QGRS in a given sequence. |
| Map | QGRS are mapped relative to locations such as splice sites in introns/exons, and poly(A) site. If original sequence was provided by user in raw or FASTA format, this step is omitted. |
| Score | QGRS are assigned scores related to their potential to form G-quadruplexes |
| Overlap Results | Overlapping QGRS are eliminated using a greedy algorithm Display summary of QGRS analysis in three forms: Gene View, Data View and Graphics View |

Figure 1. QGRS Mapper Analysis Page. This page allows the user to enter a nucleotide sequence or to search NCBI for a nucleotide entry that will then be analyzed for the presence of QGRS. The user may enter the sequence in raw or FASTA format or search NCBI by Gene ID, Gene name or Accession number. There are options to define the QGRS for the search, including: the maximum length, the minimum number of guanine tetrads, loop length and composition.

Table 3. Search Results for Bcl2. (Gene ID links to Entrez Gene)

| No. | Gene symbol | Gene ID | Organism | Analyze for QGRS |
|-----|-------------|------------------------|--------------------------|--------------------|
| 1 | Bcl2 | 12043 | <i>Mus musculus</i> | GO |
| 2 | bcl2 | 570772 | <i>Danio rerio</i> | GO |
| 3 | BCL2 | 403416 | <i>C.familiaris</i> | GO |
| 4 | BCL2 | 596 | <i>Homo sapiens</i> | GO |
| 5 | BCL2 | 281020 | <i>Bos taurus</i> | GO |
| 6 | BCL2 | 494190 | <i>Ovis aries</i> | GO |
| 7 | BCL2 | 396282 | <i>Gallus gallus</i> | GO |
| 8 | Bcl2 | 24224 | <i>Rattus norvegicus</i> | GO |
| 9 | BCL2 | 493934 | <i>Felis catus</i> | GO |

(non-overlapping and overlapping), number of QGRS found near RNA processing sites, and a visual map of the product.

As an example, the Gene View for the human GREB1 is displayed in Figure 2, showing the table of gene information and product information for the first product (the output for all products may be seen in the supplementary material).

At this stage in the analysis the user can choose among three further displays: 'Data View', 'Data View (with overlaps)' and 'Graphics View'. This can be done for the entire gene or for any particular product.

In the Data View, a table is displayed showing information for each of the set of non-overlapping QGRS. This table

displays the position of the QGRS, which exon/intron it appears in, its distance from 3' and 5' splice sites, the QGRS sequence (with each G-group underlined) and the corresponding G-score. Similar display is also shown for each QGRS mapped to poly(A) region in the product. If the user requests the Data View for the entire gene, then the QGRS information is shown for each product. The 'Data View (with overlaps)' gives the same information but shows the locations of all QGRS. Figure 3 shows the Data View for product 1 of the GREB1 gene.

The user can also choose the Graphics View to give a visual display of the location of QGRS. This allows the user to see the location of QGRS relative to exons and introns (if that information is available). The Graphics View has the following components.

- A graphic display of the entire gene (showing the location of the exons). This display includes a sliding window that can be used to focus on any particular segment of the gene. This window may be dragged to the left or right to change position within the gene.
- A magnified view of the fragment of the gene within the sliding window.
- A graph showing the location of QGRS within the fragment, with each QGRS being displayed by a bar whose height represents its G-score.
- A vertical slider that allows the user to change the size of the window. This allows the user to zoom in or out on any part of the gene. The sliding window on the gene expands or

Gene View

Search Parameters: QGRS Max Length: 30 | Min G-Group Size: 2 | Loop size: from 0 to 36 | Loop search string:

| Gene Information | |
|--|---|
| Gene ID: 9687 | Organism: Homo sapiens |
| Gene Symbol: GREB1 | Chromosome: 2 |
| Gene Name: GREB1 protein | Number of products: 3 |
| GI: 51511462 | Number of pre-mRNA poly-A Signals: 1 |
| Gene length: 108671 nt. | Aliases: KIAA0575 |
| GenBank format: NC_000002.9 | Non-overlapping QGRS: 631 |
| FASTA format: NC_000002.9 | Overlapping QGRS: 24676 |

Analyze All Products: [Data View](#) [Data View \(with overlaps\)](#) [Graphics View](#)

The gene has 3 known alternatively spliced RNA products

| Product #1: NM_014668.2 | |
|---|---|
| Number of exons: 33 | Total # of QGRS: 631 |
| Number of introns: 32 | Total # of QGRS (including overlaps): 24676 |
| Number of poly-A signals: 1 | QGRS mapped near RNA processing sites: 93 |
| Product map: 5'  3' | |

[Data View](#) [Data View \(with overlaps\)](#) [Graphics View](#)
Java Plugin Required

Figure 2. GREB1—Gene View. The first table in this view gives a summary of the gene being analyzed. This includes information such as the gene ID, name, symbol, a link to the corresponding NCBI entry, and the number of alternatively processed products and poly(A) signals found in the gene. The total number of QGRS in the gene is given for both non-overlapping and overlapping sequences. For each product, the numbers of exons and introns that were analyzed are given, together with the number of QGRS. Also, an exon/intron map is given. For the entire gene or for a particular product the user may select from three buttons to obtain more detailed information about the QGRS in the gene. These buttons are labeled Data View, Data View (with overlaps) and Graphics View. Only one product is shown in the figure. The screenshot showing all products can be found in the Supplementary Data.

QGRS sequences found (overlaps not included)

| Gene Information | |
|--------------------------------------|---|
| Gene ID: 9687 | Number of Products: 3 |
| Gene Symbol: GREB1 | Number of poly A Signals: 0 |
| Gene Size: 108671 nt. | QGRS found: 631 |
| | QGRS found (including overlaps): 24676 |

Product #1: [NM_014668.2](#) [EXONS]

| Position | Exon # | Distance from 3' Splice Site | Distance from 5' Splice Site | QGRS | G-Score |
|----------|--------|------------------------------|------------------------------|---------------------------------------|---------|
| 103 | 1 | | 10 | <u>GGAAAGGACATGGCTTTTAAACACGTGTGG</u> | 9 |
| 28386 | 3 | 38 | 67 | <u>GGAGAAGGAGGGCTGG</u> | 19 |
| 42247 | 5 | 9 | 147 | <u>GGAAATTGTGTTGGCTGTGGAAGAAAGG</u> | 16 |
| 51123 | 8 | 77 | 17 | <u>GGCTGCCCCCAAGGTGGTGGG</u> | 12 |
| 51675 | 9 | 28 | 102 | <u>GGTTGGCTTGGTGGG</u> | 20 |
| 58671 | 11 | 10 | 319 | <u>GGTGCCCTTGATGGAGGACCTGG</u> | 12 |
| 61212 | 12 | 76 | 40 | <u>GGTTACGGGGAAGG</u> | 17 |
| 63769 | 14 | 103 | 70 | <u>GGACCTGGGATCCTTTGAGAAGGTGG</u> | 10 |
| 66724 | 16 | 39 | 144 | <u>GGTGGGATTGGTGG</u> | 19 |
| 66756 | 16 | 71 | 111 | <u>GGGAGGTGAAGGAGG</u> | 18 |
| 76739 | 18 | 166 | 148 | <u>GGCCACGGGCTCATGGTCTGCTGCCG</u> | 17 |
| 76783 | 18 | 210 | 120 | <u>GGCGGTGGTGG</u> | 21 |
| 76818 | 18 | 245 | 78 | <u>GGCCCCGGCTGGCGCTGG</u> | 19 |

Figure 3. GREB1—Data View. A table of gene information is given at the top of the figure. The second table lists all QGRS mapped to exons and introns of the product, including information about the position of the QGRS, its distance from 3' and 5' splice sites, the actual sequence (underlining the G-groups) and its G-score.

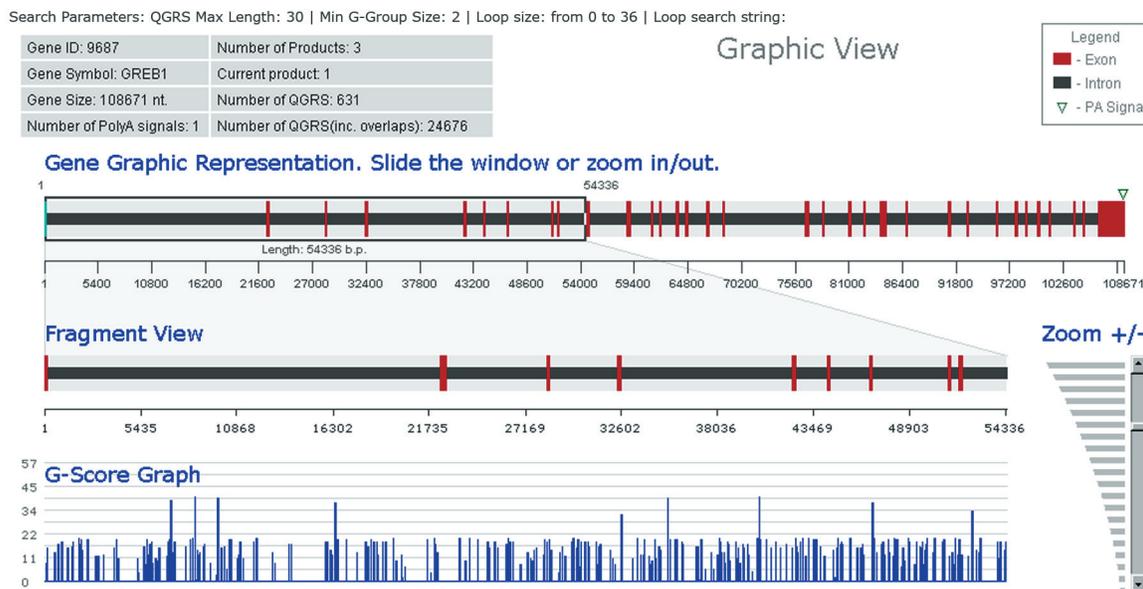


Figure 4. GREB1—Graphics View. A graphic representation of the gene, a sliding window that shows a fragment of the gene, and a graph showing the location and G-score for the non-overlapping QGRS in that fragment. A tool is provided to let the user zoom in or out on any portion of the gene. The nucleotide sequence is visible at maximum zoom levels.

contracts as one zooms in or out. It is possible to see the nucleotide sequence of the product at maximum zoom levels.

The Graphics View for the entire gene shows the G-score graph together with an exon/intron map for each product. This allows the user to visually compare the location of QGRS for each product relative to that of splice sites. The Graphics View for the first product of the GREB1 gene is represented in Figure 4. The Graphics View for the entire GREB1 gene may be seen in the Supplementary Data.

CONCLUSIONS

QGRS Mapper is a user-friendly web-based server that provides computational tools for prediction of Quadruplex forming G-rich sequences in the nucleotide sequences identified or provided by the user. The program offers many options, including user-defined composition of the quadruplex. It can analyze DNA or RNA sequence provided by the user in the raw or FASTA format. The application also provides tools for searching and retrieving gene/nucleotide entries from a variety of NCBI databases. There are several options for data output format, including an interactive graphic module.

Researchers interested in evaluating the ability of nucleotide sequences to form unimolecular G-quadruplexes will find QGRS Mapper to be very useful. Owing to the flexible and comprehensive nature of the design, it is expected to serve a variety of scientists. The application will be especially attractive to individuals interested in exploring the role of G-quadruplexes in regulated RNA processing. We are using the server to perform a large-scale analysis of alternatively processed mammalian transcripts. We are particularly interested in studying the composition and distribution patterns

of G-quadruplexes in the transcribed regions of mammalian genes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank Manuel Viotti for assistance with the testing of QGRS Mapper. This project was funded in part by a grant from the Ramapo College Foundation of Ramapo College of New Jersey. Funding to pay the Open Access publication charges for this article was provided by the Divisions of Student Affairs and Academic Affairs of Ramapo College of New Jersey.

Conflict of interest statement. None declared.

REFERENCES

- Simonsson, T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
- Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
- Davis, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem. Int. Ed. Engl.*, **43**, 668–698.
- Sharma, S., Doherty, K.M. and Brosh, R.M., Jr (2005) DNA helicases as targets for anti-cancer drugs. *Curr. Med. Chem. Anticancer Agents*, **5**, 183–199.
- Kelland, L.R. (2005) Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics—current status and future prospects. *Eur. J. Cancer*, **41**, 971–979.
- Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.

7. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macronuclei. *Proc. Natl. Acad. Sci. USA*, **98**, 8572–8577.
8. Halder, K. and Chowdhury, S. (2005) Kinetic resolution of bimolecular hybridization versus intramolecular folding in nucleic acids by surface plasmon resonance: application to G-quadruplex/duplex competition in human c-myc promoter. *Nucleic Acids Res.*, **33**, 4466–4474.
9. Xu, Y. and Sugiyama, H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.*, **34**, 949–954.
10. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
11. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chem. Biol.*, **1**, 167–173.
12. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
13. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nature Struct. Mol. Biol.*, **12**, 847–854.
14. Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl. Acad. Sci. USA*, **102**, 10864–10869.
15. Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y. and Akman, S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.
16. Shin-Ya, K. (2005) Novel antitumor and neuroprotective substances discovered by characteristic screenings based on specific molecular targets. *Biosci. Biotechnol. Biochem.*, **69**, 867–872.
17. Tahara, H., Shin-Ya, K., Seimiya, H., Yamada, H., Tsuruo, T. and Ide, T. (2005) G-quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene*, **25**, 1955–1966.
18. Zhou, J.M., Zhu, X.F., Lu, Y.J., Deng, R., Huang, Z.S., Mei, Y.P., Wang, Y., Huang, W.L., Liu, Z.C., Gu, L.Q. *et al.* (2006) Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene*, **25**, 503–511.
19. Moore, M.J., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F.A., Wilson, W.D. and Neidle, S. (2006) Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J. Med. Chem.*, **49**, 582–599.
20. Ladame, S., Schouten, J.A., Roldan, J., Redman, J.E., Neidle, S. and Balasubramanian, S. (2006) Exploring the recognition of quadruplex DNA by an engineered cys(2)-his(2) zinc finger protein. *Biochemistry*, **45**, 1393–1399.
21. Kaucher, M.S., Harrell, W.A. Jr. and Davis, J.T. (2006) A unimolecular G-quadruplex that functions as a synthetic transmembrane Na⁺ transporter. *J. Am. Chem. Soc.*, **128**, 38–39.
22. Bashkurov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M. and Heyer, W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell Biol.*, **136**, 761–773.
23. Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A.C. and Vagner, S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.*, **278**, 39330–39336.
24. Oliver, A.W., Bogdarina, I., Schroeder, E., Taylor, I.A. and Kneale, G.G. (2000) Preferential binding of fd gene 5 protein to tetraplex nucleic acid structures. *J. Mol. Biol.*, **301**, 575–584.
25. Bagga, P.S., Ford, L.P., Chen, F. and Wilusz, J. (1995) The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. *Nucleic Acids Res.*, **23**, 1625–1631.
26. Bagga, P.S., Arhin, G.K. and Wilusz, J. (1998) DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation *in vitro*. *Nucleic Acids Res.*, **26**, 5343–5350.
27. Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C. and Wilusz, J. (2002) Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res.*, **30**, 1842–1850.
28. Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J. and Milcarek, C. (2001) hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.*, **21**, 1228–1238.
29. Bruce, S.R., Dingle, R.W. and Peterson, M.L. (2003) B-cell and plasma-cell splicing differences: a potential role in regulated immunoglobulin RNA processing. *RNA*, **9**, 1264–1273.
30. Garneau, D., Revil, T., Fiset, J.F. and Chabot, B. (2005) Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.*, **280**, 22641–22650.
31. Tian, B., Hu, J., Zhang, H. and Lutz, C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
32. Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
33. Han, K., Yeo, G., An, P., Burge, C.B. and Grabowski, P.J. (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.*, **3**, e158.
34. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
35. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
36. Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L. and Bagga, P. (2006) GRSDDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D124.
37. Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L. and Hovorun, D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.*, **31**, 1375–1386.
38. Kankia, B.I., Barany, G. and Musier-Forsyth, K. (2005) Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, **33**, 4395–4403.
39. Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
40. Petraccone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Mattia, C.A., Barone, G. and Giancola, C. (2005) Relative stability of quadruplexes containing different number of G-tetrads. *Nucleosides Nucleotides Nucleic Acids*, **24**, 757–760.
41. Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
42. D'Antonio, L. and Bagga, P.S. (2004) Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences. *Proceedings of the IEEE Conference on Computational Systems Bioinformatics (CSB 2004)*. Stanford University, CA, pp. 561–562.