Nucleic Acids Research, 2006, Vol. 34, Web Server issue ***W683–W685***
doi:10.1093/nar/gkl299

Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences

Vinod Scaria, Manoj Hariharan, Amit Arora[1] and Souvik Maiti[1,*]

GN Ramachandran Knowledge Center for Genome Informatics and [1]Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

Received February 14, 2006; Revised March 10, 2006; Accepted April 6, 2006

ABSTRACT

G-quadruplex secondary structures, which play a structural role in repetitive DNA such as telomeres, may also play a functional role at other genomic locations as targetable regulatory elements which control gene expression. The recent interest in application of quadruplexes in biological systems prompted us to develop a tool for the identification and analysis of quadruplex-forming nucleotide sequences especially in the RNA. Here we present Quadfinder, an online server for prediction and bioinformatics of uni-molecular quadruplex-forming nucleotide sequences. The server is designed to be user-friendly and needs minimal intervention by the user, while providing flexibility of defining the variants of the motif. The server is freely available at URL http://miracle.igib.res.in/quadfinder/.

INTRODUCTION

Quadruplexes are higher order secondary structures formed by G-rich nucleic acid stretches in the presence of monovalent cations by Hoogstein hydrogen bonding (1). Quadruplex motifs have been known to occur in telomeres and repetitive DNA elements (2). They have gained importance in the light of discoveries unraveling their biological roles, especially as regulatory elements (3,4) and as a novel drug target against a number of pathological conditions ranging from carcinogenesis (3,5,6) to viral infections (7). Recent evidence (4) suggests that G-quadruplexes have regulatory roles in prokaryotes. The demonstration of a functional quadruplex in the promoter region of c-myc oncogene has illustrated the potential therapeutic importance of such structures.

The double-stranded nature of DNA implies that quadruplexes formed by G-rich strand must exist in competition with the normal double-stranded Watson–Crick paired duplex. However, RNA being single-stranded, is unaffected by such competition and therefore functional role for G-quadruplexes hold more promise at the RNA level. These motifs have already been implicated in a variety of biological processes at RNA level like translation initiation (8), repression (9) and are thought to play an important role in pathophysiological processes like Fragile X mental retardation by virtue of interaction with FMRP (10). Moreover recent evidence implicates G-quadruplexes in tissue specific alternative splicing events (11).

The full spectrum of diverse biological roles of quadruplex-forming sequences is slowly being unraveled with the identification of proteins and ligands (12) which recognize quadruplex motifs as well as factors which influence the equilibrium (13) of quadruplex motifs.

Apart from their diverse biological roles, quadruplex-forming aptamers have been used recently for designing molecular sensors (14), synthetic ion-channels (15) and molecular motors (16). Clearly, the applications of quadruplex-forming sequences are still emerging.

The study of the wide spectrum of functions and processes which involve quadruplexes will benefit from the availability of a tool which predicts quadruplex-forming DNA/RNA sequences. Though a couple of papers have appeared recently on genome-wide analyses of quadruplex-forming sequences encoded by the human genome (17,18), to the best of our knowledge, this is the first server available for prediction and analysis of quadruplex motifs.

WEB APPLICATION

Quadruplex-forming motif

We employ a consensus (17,18) uni-molecular G-quadruplex sequence motif of the form $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x denotes the G-stretch and $y1$, $y2$ and $y3$ denote the loop lengths. We search for all possible motifs, including overlapping ones by brute force. The algorithm runs for user-defined variables and there is no restriction to the sequence length or the variables.

*To whom correspondence should be addressed. Tel: +91 011 27666156; Fax: +91 011 27667471; Email: souvik@igib.res.in

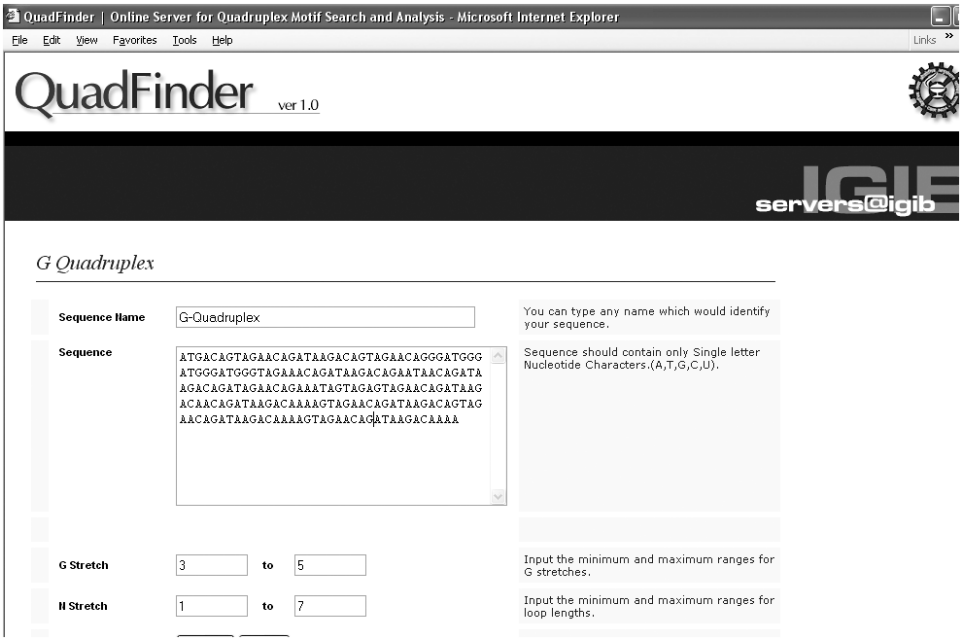

Figure 1. Input options for Quadfinder. The user is asked to input the nucleotide sequences along with the minimum and maximum values of G-stretches and loop lengths.

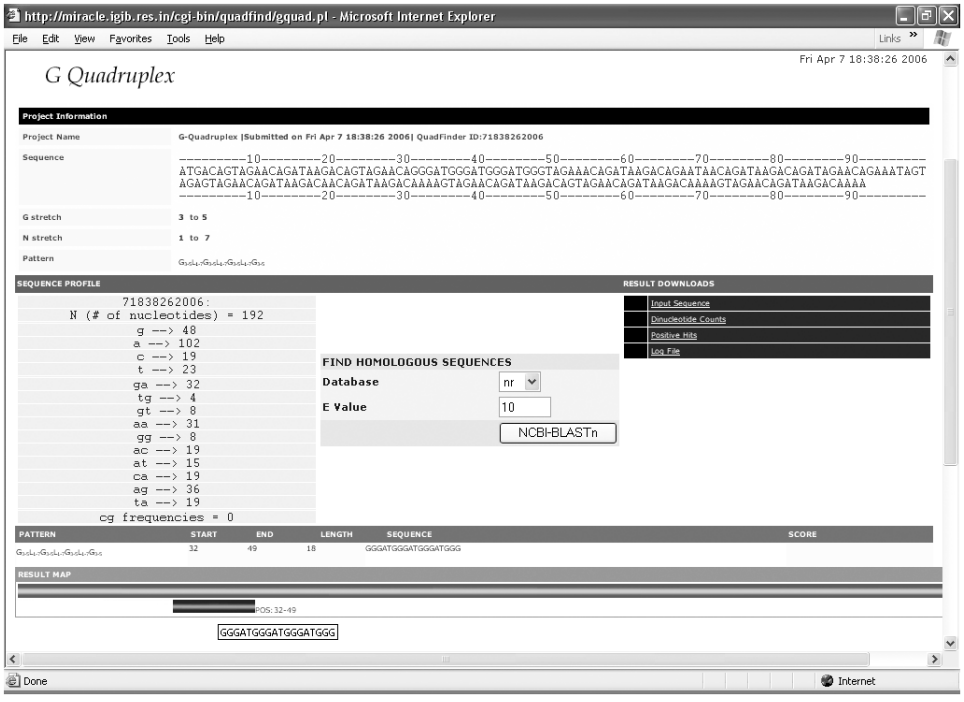

Figure 2. Output display of Quadfinder. This page displays the positive hits along with options to further analyze the motif. The user also has the convenience to download the result files.

Implementation and interfaces

The server is implemented in CGI/Perl and runs on Apache HTTP server version 2.0. The server interface is designed to be user-friendly and takes minimum user inputs (Figure 1). The inputs include the nucleotide sequence to query and the maximum and minimum lengths of the G-stretches and the loop lengths. Though earlier attempts at genome-wide search of G-quadruplex motifs have restrained the loop lengths citing computational complexities (18), we have implemented a more flexible search option whereby the user has the convenience to set the parameters, even

while it searches for the default motif ($3 \leq x \leq 5; 1 \leq y \leq 7$), where x denotes the G-stretch and y denotes the loop lengths.

The server displays the hits both in tabular form and with a diagrammatic representation mapping hits back into the sequence. The user also has an option to download the predictions at a later point of time through unique submission IDs. In addition to providing information on potential quadruplex-forming motifs, important information on the nucleotide sequence features like di-nucleotide frequencies are also provided (Figure 2). The user also has the convenience to search for homologous sequences using the BLAST interface.

In the near future, the server will be highly interconnected to other biological databases providing the user the flexibility of using gene identifiers instead of sequences as input. We also plan to provide pre-computed datasets for eukaryotic genomes, which would make it a comprehensive suite for the computational analysis of quadruplex motifs.

DISCUSSION AND CONCLUSIONS

Quadfinder is a tool for search and analysis of quadruplex-forming motifs in nucleotide sequences. The tool enables users to discover G-quadruplex motifs in any sequence of interest. The server is designed to be user-friendly so that researchers with minimal computational skills can use it. The diagrammatic representations of results facilitate better understanding of the spatial orientation of the motifs with respect to the input sequence. Moreover, the scoring of quadruplex motifs enables to prioritize motifs for further experimental studies. In addition, the user has an option to retrieve the results of an earlier analysis at a later point of time making it a unique analysis suite for quadruplex motifs.

ACKNOWLEDGEMENTS

The authors thank Dr Beena Pillai for reviewing the manuscript, and anonymous reviewers for suggesting improve ments. The authors would also like to acknowledge the Council for Scientific and Industrial Research (CSIR), India for funding through CMM0017. V.S. and A.A. are recipients of Research Fellowship from CSIR and University Grants Commission, Goverment of India, respectively. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
2. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
3. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *PNAS*, **99**, 11593–11598.
4. Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: Role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
5. 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.
6. Dai,J., Dexheimer,T.S., Chen,D., Carver,M., Ambrus,A., Jones,R.A. and Yang,D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **128**, 1096–1098.
7. 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.
8. 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.
9. 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.
10. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
11. Kostadinov,R., Malhotra,N., Viotti,M., Shine,R., D'Antonio,L. and Bagga,P. (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D124.
12. Maiti,S., Chaudhury,N.K. and Chowdhury,S. (2003) Hoechst 33258 binds to G-quadruplex in the promoter region of human c-myc. *Biochem. Biophys. Res. Commun.*, **310**, 505–512.
13. Kumar,N. and Maiti,S. (2005) The effect of osmolytes and small molecule on Quadruplex-WC duplex equilibrium: a fluorescence resonance energy transfer study. *Nucleic Acids Res.*, **33**, 6723–6732.
14. Radi,A.E., AceroSanchez,J.L., Baldrich,E. and O'Sullivan,C.K. (2006) Reagentless, reusable, ultrasensitive electrochemical molecular beacon aptasensor. *J. Am. Chem. Soc.*, **128**, 117–124.
15. Kaucher,M.S., Harrell,W.A. and Davis,J.T. (2006) A unimolecular G-quadruplex that functions as a synthetic transmembrane Na+ transporter. *J. Am. Chem. Soc.*, **128**, 38–39.
16. Wang,Y., Zhang,Y. and Ong,N.P. (2005) Speeding up a single-molecule DNA device with a simple catalyst. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **72**, 051918.
17. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
18. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.