Bioinformatics, 36(10), 2020, 3246–3247
doi: 10.1093/bioinformatics/btaa057
Advance Access Publication Date: 27 January 2020
Applications Note

Sequence analysis

G4Killer web application: a tool to design G-quadruplex mutations

Vaclav Brazda[1], Jan Kolomaznik[2], Jean-Louis Mergny[1,*] and Jiri Stastny[2,*]

[1]Institute of Biophysics of the Czech Academy of Sciences, Brno 612 65, Czech Republic and [2]Department of Informatics, Mendel University in Brno, Brno 613 00, Czech Republic

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on November 12, 2019; revised on December 20, 2019; editorial decision on January 19, 2020; accepted on January 21, 2020

Abstract

Motivation: G-quadruplexes (G4) are important regulatory non-B DNA structures with therapeutic potential. A tool for rational design of mutations leading to decreased propensity for G4 formation should be useful in studying G4 functions. Although tools exist for G4 prediction, no easily accessible tool for the rational design of G4 mutations has been available.

Results: We developed a web-based tool termed *G4Killer* that is based on the G4Hunter algorithm. This new tool is a platform-independent and user-friendly application to design mutations crippling G4 propensity in a parsimonious way (i.e., keeping the primary sequence as close as possible to the original one). The tool is integrated into our DNA analyzer server and allows for generating mutated DNA sequences having the desired lowered G4Hunter score with minimal mutation steps.

Availability and implementation: The G4Killer web tool can be accessed at: http://bioinformatics.ibp.cz.

Contact: stastny@fme.vutbr.cz or mergny@ibp.cz

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Recently, an increasing number of articles have pointed to important biological roles of local non-B DNA structures. Several algorithms are accessible to analyze such various local DNA structures as cruciforms, triplexes, and G-quadruplexes (G4). Increasing efforts are being devoted to the study of G4 due to their possible involvement in serious pathologies and great stability under physiological conditions. G4 play important roles in transcription, translation and RNA, genomic stability, telomere biology, protein recognition and replication origin definition.

G4-forming sequences are common in many regulatory regions of the human genome (Huppert and Balasubramanian, 2007). To evaluate the role of a particular G4-forming sequence in the genome, targeted mutations that do not change length and minimally perturb the primary sequence are required. We therefore designed the 'G4Killer' web application, a new and easily accessible tool for the evaluation and mutation of G4-forming sequences. Our algorithm allows defining conditions for G4Hunter score requirements and determining the minimal number of point mutations abolishing G4 propensity. This platform-independent tool has an easy-to-use, web-based graphical interface, freely available at http://bioinformatics.ibp.cz : 8888/#/analyse/g4-killer.

2 Features

G4Killer is an application with a web interface. A user inserts either a single sequence or multiple sequences (up to 10, with an upper limit for sequence length of 200 bp) separated by a newline character. The next step is to specify a maximum G4Hunter score (G4HS; default value = 1) so that the proposed mutant sequence has a G4HS equal to or lower than this threshold. The result is a sequence or sequences with the minimal number of mutations and lowest G4Hunter score achievable below the threshold. The number of sequences depends on the input. If the input sequence contains multiple G-tracks of identical length and/or longer repeats of Gs, then multiple equivalent results with mutations in different bases but having the same score are possible and displayed. This effect results from the way G4HS is calculated. For example, a GGGG run can be mutated to a single G and a double G by changing either the second or third position to A or T (the resulting GAGG, GTGG, GGAG or GGTG motifs will have the same G4HS). G=>C mutations that would further decrease G4HS is not considered here in order to avoid GC-rich strands that may form stable competing hairpins. The user can see all possible solutions in single sequence mode, while only one possible solution is displayed in multiple sequence mode. The sequence with the lowest score is shown; if there are multiple

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/article/36/10/3246/5716328 by VUT Brno user on 26 August 2020

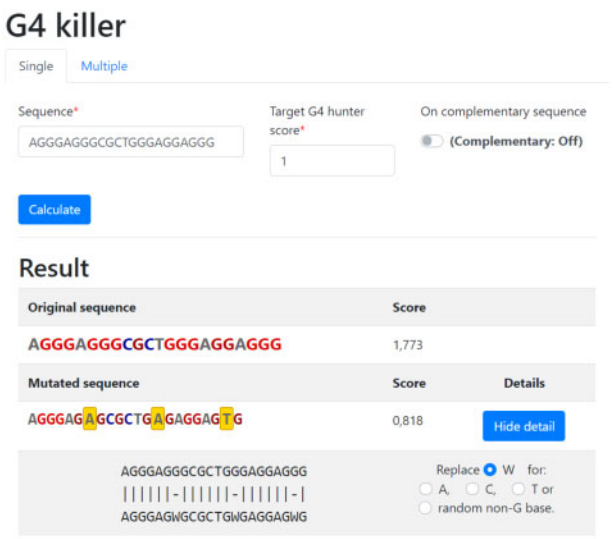

Fig. 1. G4Killer result screen wherein the G-rich motif found in the c-kit promoter is targeted. In the upper part is a form for the user's data input (sequence, target G4Hunter score, switch for analyses of C-rich DNA sequences). After calculation, the Result part is shown containing original and mutated sequences with their G4Hunter scores. G-tracks are shown in red (the brighter the display, the more Gs in the track). In the Details part, alignments of the original and mutated sequences are shown. (Color version of this figure is available at *Bioinformatics* online.)

equivalent sequences, one is selected randomly. For C-rich sequences with negative G4Hunter scores, there is an option to perform calculations with the complementary sequence where the input is transformed into complementary sequence and then standard calculations are performed. The mutations required for reducing the G4Hunter score are highlighted and the calculated G4Hunter score is presented in the Result section. The user also can apply custom rules for mutated bases replacement in the details view (Fig. 1). More examples are shown in Supplementary Material S1.

3 Validation

The algorithm itself produces only candidate sequences with mutated bases that are then evaluated using our rewrite of G4Hunter published and validated earlier (Brázda *et al.*, 2019). The algorithm is based on a breadth-first search graph algorithm. It takes the original sequence, changes each individual G-base into an A or T and calculates the score of mutated sequences. The result of this step is N copies of the original sequence with one mutated base, where N is the number of G bases. These mutated sequences are stored, and the same approach is repeated for each of them. This is repeated until the score is lower than or equal to the specified target (e.g. 1.0). The set of sequences from the last step is presented as a result. Sequences are ordered by their scores. Although it has greater memory requirements, using a breadth-first search algorithm ensures finding the best solution to the problem with the least number of steps.

4 Conclusions

We developed a web tool, the G4Killer application, for finding the minimal mutation of G4-forming sequences. Our webserver allows quick and platform-independent usage for single and multiple G4-forming sequences, including their visualization. This tool is also integrated into the G4Hunter server application. That means sequences found by the G4Hunter server can be analyzed directly by G4Killer with just one click and no requirement to copy and edit.

Acknowledgements

The authors thank Martin Bartas for testing and Gale A. Kirking for proofreading and editing the article.

Funding

This work was supported by the Czech Science Foundation [18-15548S] and SYMBIT project reg. no. [CZ.02.1.01/0.0/0.0/15_003/0000477] financed from the ERDF.

Conflict of Interest: none declared.

References

Brázda,V. *et al.* (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, **35**, 3493–3495.

Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.

Lam,E. *et al.* (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.

Downloaded from https://academic.oup.com/bioinformatics/article/36/10/3246/5716328 by VUT Brno user on 26 August 2020