

# How bioinformatics resources work with G4 RNAs

Joanna Miskiewicz, Joanna Sarzynska and Marta Szachniuk

Corresponding author: Marta Szachniuk, Institute of Computing Science, Poznan University of Technology, ul. Piotrowo 2, 60-965 Poznan, Poland. Tel.: +48 61 6653030; Fax: +48 61 8771525. E-mail: mszachniuk@cs.put.poznan.pl

## Abstract

Quadruplexes (G4s) are of interest, which increases with the number of identified G4 structures and knowledge about their biomedical potential. These unique motifs form in many organisms, including humans, where their appearance correlates with various diseases. Scientists store and analyze quadruplexes using recently developed bioinformatic tools—many of them focused on DNA structures. With an expanding collection of G4 RNAs, we check how existing tools deal with them. We review all available bioinformatics resources dedicated to quadruplexes and examine their usefulness in G4 RNA analysis. We distinguish the following subsets of resources: databases, tools to predict putative quadruplex sequences, tools to predict secondary structure with quadruplexes and tools to analyze and visualize quadruplex structures. We share the results obtained from processing specially created RNA datasets with these tools.

Contact: [mszachniuk@cs.put.poznan.pl](mailto:mszachniuk@cs.put.poznan.pl)

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

Key words: quadruplexes; RNA; databases; bioinformatics tools; structure analysis; PQS prediction

## Introduction

G-quadruplexes (G4s) are non-canonical, four-stranded structures that form in guanine-rich nucleic acids. The basic structural unit of G4 is a G-tetrad: four guanines in the planar arrangement that interact with one another via hydrogen bonds. Quadruplexes can also contain non-G based quartets; about 20% of tetrads found in PDB-deposited quadruplex structures contain other nucleotides than guanine [1, 2]: U-tetrads, mixed ATAT, GCGC tetrads, etc. [2–6]. If  $n$  ( $n \geq 2$ ) neighboring tetrads stack with one another, they create an  $n$ -layer quadruplex. Two or more G4s can associate through stacking interactions between their outer tetrads to form higher-order multimers, ranging from dimers to G-wires [7, 8]. All these structural motifs have several topological characteristics, the most important

of which are strand orientation in the stem, the number of tetrads, loops' arrangement and groove dimension. One, two or four individual strands—linked by loops—participate in the quadruplex formation. They have parallel, antiparallel or hybrid orientation [9], which correlates to anti/syn conformation of guanines in the tetrad planes, base stacking geometry and loop types [10–12].

Scientists observed the very first biologically relevant G4s in eukaryotic chromosomal telomeric DNA. This discovery initiated the research into the role and distribution of quadruplexes in the genome. Chemical biology methods were developed to map G4 folding *in vitro*, like G4-seq [13], and *in vivo*, like G4ChIP-seq [14], in the genomes of various species, including humans [15],

---

**Joanna Miskiewicz** is a PhD student and a member of Laboratory of RNA Structural Bioinformatics, Poznan University of Technology. Her research interests include structural bioinformatics, motif identification and algorithms for RNA biology. She holds an MSc in bioinformatics (2015).

**Joanna Sarzynska** is a research associate at IBCh PAS. Her research interests include RNA structure, molecular dynamics and structural bioinformatics. She has authored papers published in top scientific journals on life sciences and holds a PhD in chemistry (1997).

**Marta Szachniuk** is a professor of technical sciences, vice-president of the Polish Bioinformatics Society and vice chair of EURO CBBM. Her research interests include algorithms for structural biology, operations research and AI. She is an author of highly cited papers and over 20 bioinformatics tools. She holds a PhD (2005) and a DSc (2015) in computing science, ProFit (2020).

**Submitted:** 25 June 2020; **Received (in revised form):** 3 August 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)  
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](mailto:journals.permissions@oup.com)

[16]. First genomic maps of DNA G4s appeared—a starting point for clarifying the purpose of quadruplex formation [17]. Many sequences with G4 potential turned out to be associated with cancer or neurodegenerative diseases and became an attractive target for molecular therapeutics [18]. Extensive research on the DNA quadruplexes generated interest in RNA G4s. RNA G-quadruplex sequencing protocols—rG4-seq [19, 20], G4RP-seq [21]—were developed for detecting RNA quadruplexes on a transcriptome-wide level. The studies revealed RNA G4s in coding and noncoding fragments of mRNA [22], mature human miRNA [23], other premature and mature noncoding RNAs (including telomeric RNA) and aptamers [24]. RNA quadruplexes showed to regulate pre-mRNA processing, control RNA localization and mechanisms like mRNA translation, miRNA biogenesis or protein binding [25]. G4 in 5'UTR of mRNAs proved the ability to repress translation *in vitro* [26]. Experiments indicated that under certain conditions—the presence of salts or ligands—pre-miRNA sequences could adopt the G4 structure avoiding the Dicer-mediated maturation and leading to subexpression of miRNA levels [27]. A structural study led to the determination of first quadruplex folds via X-ray [28] and nuclear magnetic resonance (NMR) [29]. Currently, Protein Data Bank collects about 300 G4-rich nucleic acid structures, 75% of them from DNAs [2, 30, 31]. Their analysis showed differences between DNA and RNA quadruplexes. DNA G4s are structurally polymorphic and thermodynamically less stable than their RNA equivalents [1, 32]. RNA quadruplexes are predominantly parallel; ribose favors anti-oriented guanines which results in the strands' parallelism.

The research into quadruplexes resulted in an increased demand for computer resources dedicated to these motifs. Bioinformatics responded with tools for DNA G4s that pretended to process quadruplexes regardless of the molecule type. They aimed to identify G4 forming sequences, model and analyze the secondary and tertiary structures, simulate molecular dynamics, calculate free energy or perform molecular docking [33]. Many resources for predicting G4 formation and stability, available before 2010, were reviewed in [34]. With the increasing number of experimental data on quadruplexes, including more G4-rich high-resolution structures and genomic maps of DNA G4s, new algorithms appeared [15, 16]. They go beyond canonical G4 predictions and apply non-trivial computational techniques, including machine learning (ML) [35–37]. Lombardi and Londoño-Vallejo [33] present a comprehensive overview of modern open-source software for G-quadruplex detection tested on a set of G4 DNAs verified experimentally *in vitro*.

In this paper, we focused on RNA quadruplexes. We described 35 bioinformatics resources that addressed quadruplexes and checked how they worked with RNA G4s. The set included 14 tools from [33] and 21 others. We grouped all programs into four categories: (i) databases, (ii) sequence-based tools predicting G4 location in the sequence, (iii) secondary structure prediction tools and (iv) secondary and tertiary structure-based tools analyzing and visualizing quadruplexes. We performed their tests on specially created datasets. We hope that our review will be helpful to scientists studying the specifics of RNA G4s.

## Methods

In preparing this review, we searched global resources to create a potentially complete list of databases and analytical tools used in quadruplex research. As a result, the list includes 16 data repositories, 14 tools that predict quadruplex-forming sites in nucleic acid sequence, 1 tool that predicts the secondary structure and annotates potential location of quadruplexes and 4

tools that analyze the secondary and tertiary structure and visualize quadruplex topology (Figure 1). Twenty-one resources with web interfaces were tested by us mostly via a web browser. The remaining ones were downloaded, configured and run locally. In most cases, we applied the default input settings.

## Databases with G4-related data

Currently, there exist 16 databases, which store information concerning quadruplexes. They fall into three categories: databases that collect primary or tertiary structures with experimentally verified G4s (DSSR-G4DB, G4IPDB, G4LDB, G4RNA, Lit392 and Lit638); databases storing data from high-throughput sequencing with mapped quadruplexes (GSE63874, GSE77282, GSE110582 and GSE129281); and databases of sequences with G4s identified *in silico* (Greglist, GRSDDB2, G4-virus, Non-B DB v2.0, Plant-GQ and QuadBase2). We describe them briefly in the following paragraphs and define the following features in Table 1: DNA and RNA indicate whether the database collects DNA and RNA sequences; G4 verification denotes whether quadruplexes are verified experimentally or predicted *in silico*; G4 sequence informs if the quadruplex sequence is available; the number of G4s gives the number of stored quadruplex sequences (as of 21 March 2020); DB records specifies the number and type of database entries (as of 21 March 2020); customized search shows whether the database has a search engine that allows to search its records with different criteria; web interface specifies if the database is web-interfaced; visual output indicates whether any visualization of the output data is available.

DSSR-G4DB [38] contains quadruplex nucleic acid structures found by DSSR in the Protein Data Bank [30], currently 354 entries. The data are annotated. Users can find information about G-tetrads, G4 helices and G4-stems and visualize the 3D models of G4 structures. Availability: webserver (<http://g4.x3dna.org>). Recent update: 5 June 2020.

G4IPDB [39] is a database of over 200 proteins interacting with DNA and RNA G-quadruplexes, based on the literature data. For each entry, it contains the G4 sequence, interacting protein name, and UniProt ID, the details of the interaction, PubMed ID of the paper being the source of information. Users browse the data and query the database by specifying G4IPDB interaction ID, DNA/RNA target name, gene name, etc. Availability: webserver (<http://bsbe.iiti.ac.in/bsbe/ipdb>). Updated: twice a year.

G4LDB [40] collects ligands (currently, over 800) that interact with G-quadruplexes. Each entry contains information about chemical structure, targeted G4 sequence, physical properties of a ligand and literature references. The 3D model of every ligand is also available. Users browse the database and search for ligands by defining their structure, ligand properties, ligand activity fields or bibliographic information. Availability: webserver (<http://www.g4ldb.org/ci2/index.php>). Updated: twice a year.

G4RNA [36] stores published human RNA sequences, processed experimentally. This collection includes sequences with confirmed G-quadruplexes and sequences confirmed not to form G4s, currently, 567 entries in total. The system allows running a keyword-driven and position-driven search. Keywords include the G4 gene symbol or sequence (IUPAC-encoded or regular expression), the experiment name, reference paper and DOI. Position-driven search needs specifying at least one chromosome to contain G4. The results fall into four categories: sequences (nine options), experiments (four options), predictions (seven options) and QGRS Mapper (nine options). Users can choose to display the secondary structure

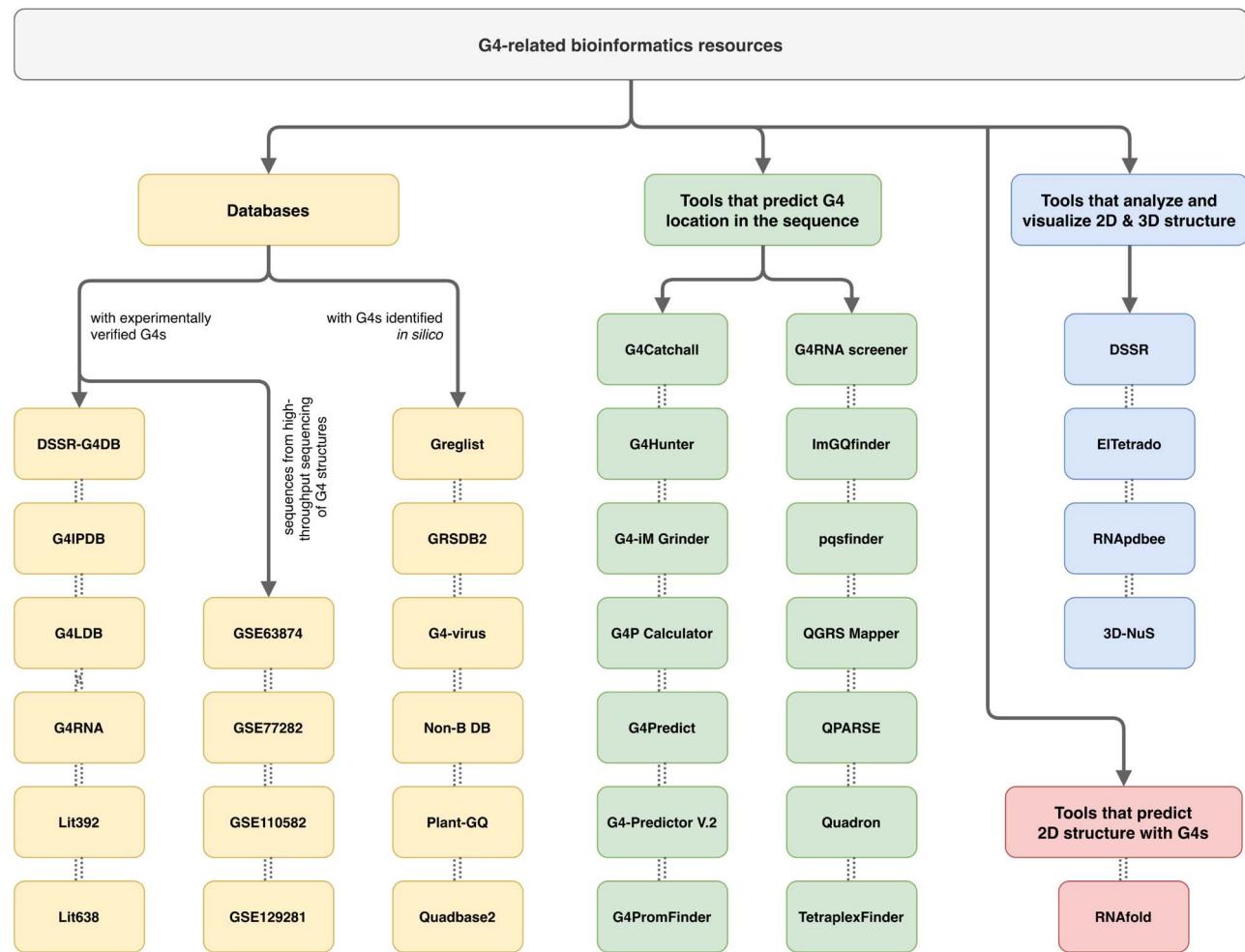


Figure 1. Bioinformatics resources for quadruplexes.

predicted by RNAfold with quadruplex annotated in the dot-bracket representation. The output data can be download in the xls file. Availability: webserver (<http://scottgroup.med.usherbrooke.ca/G4RNA>). Updated: monthly.

Lit392 [41] is a set of 392 DNA and RNA sequences for which the formation of G4s was experimentally confirmed (298 sequences) or disproven (94 sequences). The set mainly includes published sequences and several unpublished ones resulting from the experiments performed by the authors. The database was created to test the performance of G4Hunter. Availability: supplementary file to [41] ([https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4770238/bin/supp\\_gkw006\\_nar-02998-f-2015-File003.zip](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4770238/bin/supp_gkw006_nar-02998-f-2015-File003.zip)). Updated: no.

Lit638 [42] combines data from Lit392 and G4RNA databases. It contains 638 DNA and RNA sequences: 506 confirmed to form G4s and 132 non-forming quadruplexes. The database was created to test the performance of QPARSE method. Availability: upon request to authors. Updated: no.

GSE63874 [15] is a map of distinct canonical and non-canonical G4s, experimentally confirmed to form in the human genome. It consists of 716,310 DNA G4s, obtained from the high-throughput G4-seq method. The genomic DNA template was sequenced twice, first with Na<sup>+</sup> and second with K<sup>+</sup>. Binding to K<sup>+</sup> cations enhances the structural stability of G4. Availability: BED files on GEO webserver (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63874>). Updated: no.

GSE77282 [19] is a map of canonical and non-canonical G4s found in the human transcriptome. RNA G4s were identified in vitro by the rG4-seq method—over 3000 in the presence of K<sup>+</sup> cation or the PDS ligand. Availability: BED files on GEO webserver (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77282>). Updated: no.

GSE110582 [16] is a map of DNA G4s, identified in the genomes of 12 species (model organisms—including human—and clinically important pathogens). The data—1,420,841 G-quadruplexes—come from high-throughput sequencing applying G4-seq2 (improved G4-seq method) with K<sup>+</sup>, Li<sup>+</sup> and PDS ligand as the sequencing buffer. Availability: BED files on GEO webserver (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110582>). Updated: no.

GSE129281 [43] is a map of G4 structures found in the genomes of *Pseudomonas aeruginosa* and *Escherichia coli*. The rG4-seq method, in the presence of K<sup>+</sup> or Li<sup>+</sup> cations, revealed 329 RNA G4 sites—168 in *E. coli* and 161 in *P. aeruginosa*. Availability: BED files on GEO webserver (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129281>). Updated: no.

Greglist [44] is a database of potential G-quadruplex regulated genes in the genomes of various species, including humans. Users browse through the database or search for stored DNA putative quadruplex sequences (PQSs). Database records contain, among others, Ensembl ID, gene name, organism, number of PQSs, G4 sequence and distance to TSS. PQSs were

Table 1. Selected features of G4-related databases

Database	DNA	RNA	G4 verification	G4 sequence	Number of G4s	DB records	Customized search	Web interface	Visual output
DSSR-G4DB	✓	✓	experimental	✓	354	354 (PDB structures)	✓	✓	✓
G4PDB	✓	✓	experimental	✓	no data	216 (interactions)	✓	✓	✓
G4LDB	✓	✓	experimental	✓	no data	> 800 (ligands)	✓	✓	✓
G4RNA	✓	✓	experimental	✓	321	567 (human RNA sequences)	✓	✓	✓
Lit392	✓	✓	experimental	✓	298	392 (DNA and RNA sequences)	✓	✓	✓
Lit638	✓	✓	experimental	✓	506	638 (DNA and RNA sequences)	✓	✓	✓
GSE63874	✓	✓	experimental	✓	716,310	32 million (reads)	✓	✓	✓
GSE77282	✓	✓	experimental	✓	3383	1.15 billion (reads)	✓	✓	✓
GSE110582	✓	✓	experimental	✓	1,420,841	7,675,39 Mb	✓	✓	✓
GSE129281	✓	✓	experimental	✓	329	3505 (hits)	✓	✓	✓
Greglist	✓	✓	in silico	✓	no data	115442 (genes)	✓	✓	✓
GRSDB2	✓	✓	in silico	✓	3,255,075	29,288 (genes)	✓	✓	✓
G4-virus	✓	✓	in silico	✓	47	248 (viruses)	✓	✓	✓
Non-B DB v2.0	✓	✓	in silico	✓	3,864,596	12 (mammalian genomes)	✓	✓	✓
Plant-GQ	✓	✓	in silico	✓	626,341,645	195 (plants)	✓	✓	✓
QuadBase2	✓	✓	in silico	✓	no data	1897 (species)	✓	✓	✓

found using Quadparser [52] with  $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$  and  $C_{3+}N_{1-7}C_{3+}N_{1-7}C_{3+}N_{1-7}C_{3+}$  (to find quadruplexes on complementary strand). Availability: webserver (<http://tubic.tju.edu.cn/greglist>). Recent update: 19 October 2007.

GRSDB2 [45] stores PQS positions in computationally processed over 29,000 eukaryotic pre-mRNA sequences of eight organisms, including humans. QGRS Mapper, run to search for  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$  motif, identified 3,015,683 PQS for  $x = 2$  and motif length  $\leq 30$  nts, and 239,392 PQS for  $x = 3$  and motif length  $\leq 45$  nts. Availability: webserver (<http://bioinformatics.ramapo.edu/GRSDB2/>). Updated: no.

G4-virus [46] is a collection of PQS locations in the human viruses' genomes. Users search viruses by name or browse the database by virus class. Deposited data consist of PQS identified computationally on both strands, PQS positions in viral genomes, conservation degrees among different strains and statistical data for PQS. The system provides a graphical visualization of the PQS arrangement. Availability: webserver ([http://www.medcomp.medicina.unipd.it/main\\_site/doku.php?id=g4virus](http://www.medcomp.medicina.unipd.it/main_site/doku.php?id=g4virus)). Recent update: 30 July 2019.

Non-B DB v2.0 [47] contains non-B DNA structure predictions of 3,864,596 quadruplex forming motifs in 12 mammalian genomes, including the human genome. The database system allows searching by features or attributes. In plain feature search, users select species, classes, chromosome or gene type, start and stop position of the chromosome, query type (specifying how many features and where they should be found) and feature type (for example, quadruplex). Attribute search offers additional options: composition, sequence and tracts. Users can submit their data of non-B DNA motifs. The visualization of non-B DNA motifs for each organism is provided. Availability: webserver (<https://nonb-abcc.nciccrf.gov/apps/site/default>). Recent update: 13 June 2012.

Plant-GQ [48] collects 626,341,645 DNA PQS from 195 plant species: 610,897,949 of them are two-tetrad G4s; 14,326,347 are three-tetrad G4s; 1,117,349 are four-tetrad G4s. The data were obtained by searching for motifs matching the following pattern  $G_xN_yG_xN_yG_xN_yG_x$ , where  $x \in [2, 4]$ ,  $y \in [1, 10]$ . Database entries include PQS sequences and positions within the genome. Availability: webserver (<http://biodb.sdau.edu.cn/plantgq/index.php>). Updated: no.

QuadBase2 [49, 50] is a database of DNA PQS found in 178 species of eukaryotes (EuQuad module) and 1719 species of prokaryotes (ProQuad module). The data result from searching for  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$  pattern. Users introduce the query by selecting a group of organisms, the species, gene ID, region of interest (gene body, around TSS, CDS and UTRs, strand), PQS distance from TSS, search algorithm (greedy/non-greedy), bulge size, etc. Additional search parameters for G4s concern the stringency level, which can be low (two G-tetrads, loop size 1–12), medium (three G-tetrads, loop size 1–7) or high (three G-tetrads, loop size 1–3). Resulting PQSs are shown in circular histograms. Availability: webserver (<http://quadbase.igib.res.in/>). Updated: no.

### Tools that predict G4 location in the sequence

The first bioinformatics tools, searching for the existence of PQS motifs, were published in 2004–2005 [51, 52]. To this day, several such programs appeared. They differ in the search algorithm, the searched pattern and the scoring function; for example, some tools apply an ML approach, others look for sequential motifs that match different regular expressions. In Table 2, we show these tools' options important from the user point of view.

Columns in the table: DNA and RNA indicate whether the tool accepts DNA and RNA sequences; multiple entry checked means that the program allows entering many sequences for a single run; allow mismatches denotes if the tool accepts mismatches and bulges in a G-tract/PQS; number of tetrads in G4 informs about limit for the number of G-tetrads per quadruplex; loop length specifies the limit imposed on every loop involved in quadruplex formation; max PQS' length is the program restriction for PQS length; accept overlapping indicates whether the tool searches for overlapping PQS; non-G PQS informs about the possibility to search for non-G-based quadruplexes; strands tells whether the tool provides the results only for the input sequence (+) or for both—the input and the complementary strand (+/-); show G4 position informs if the tool gives the PQS location within the sequence. Different search patterns with examples are presented in the Supplementary Materials (Table S5).

G4Catchall [53] looks for conservative G4s by fitting into the regular expression and user-defined parameters. Sequences that lack typical, uninterrupted G4s (pattern:  $G_n$ , where  $n$  denotes the number of guanines) are scanned with more complicated patterns:  $G_xN_yG_{(n-x)}$  or  $G_xN_yG_{(n-x-1)}$ , where  $x \in [1, n]$  (a detailed explanation of search patterns can be found in the [Supplementary Materials](#)). A definition of the G4 motif includes bulged G-tract, mismatched tract that contains non-guanine nucleotides and loops between tracts—one extreme loop, up to 50 nucleotides, is allowed. Users can provide minimum G-tract length (2 or 3), loop length (1–15), extreme loop length (1–50), the number of allowed bulges and mismatches (0–2), complementary strand search, overlapping G4s merge and flanking nucleotides inclusion. Input data formats: raw sequence, FASTA. Availability: Python script, web application (<http://homes.ieu.edu.tr/odoluca/G4Catchall>).

G4Hunter [41, 54] predicts PQS in DNA or RNA sequence based on its G-richness and G-skewness, meaning the fraction of Gs in the sequence and G/C ratio between the strands, respectively. The program allows users to define the size of a search window (web default: 25) and the threshold value (web default: 1.2). In the input sequence, each nucleotide and nucleotide aggregation has arbitrary assigned value: score(G) = 1, score(GG) = 2, score(GGG) = 3 and score(GGGG...) = 4. Cytosines have the opposite values: score(C) = -1, score(CC) = -2, score(CCC) = -3, and score(CCCC...) = -4. Other nucleotides are not counted. Input data format: FASTA. Availability: Python script, web application (<http://bioinformatics.ibp.cz:8888/#/analyse/quadruplex>). Web application stores users results in a relational database.

G4-iM Grinder [55] finds and characterizes quadruplexes and i-Motifs within a given DNA or RNA sequence. It uses a two-part search engine with 13 customizable functions (e.g. showing PQS on both strands, loop sequence and size). The first search method (M1A) identifies uninterrupted and bulged nucleotide runs. Users define the length of the run, the size of the bulge and the nucleotide the run is composed of (any nucleobase can be used to form a run). Discovered runs are passed to the second search method (M1B). The method finds a correlation between runs, starting with closest runs in a sequence and broadening the search if necessary. The distance between the runs is limited by the users. The search results are further analyzed by methods, one for overlapping and size-dependent PQS search (M2) and second for non-overlapping size-independent search (M3). M2 links the runs for final structure depending on user-defined limitations for the number of connected runs, the length of the final motif sequence and the number of bulges within the sequence. The linkage process is followed by the frequency count of the final structure within the input sequence. M3 searches for higher-order structures based on unlinked runs from the M1B

Table 2. Selected features of PQS prediction tools

Tool	DNA	RNA	Multiple entry	Allow mismatches	Number of tetrads in G4	Loop length	Max PQS'	Accept overlapping	Non-G PQS	Strands	Show G4 position
TetraplexFinder	/	/	/	/	/	/	/	/	/	/	/
QARSE	/	/	/	/	/	/	/	/	/	/	/
QPSAS Mapper	/	/	/	/	/	/	/	/	/	/	/
QCatchall	/	/	/	/	/	/	/	/	/	/	/
QCatcher	/	/	/	/	/	/	/	/	/	/	/
QPSFinder	/	/	/	/	/	/	/	/	/	/	/
G4Predict	/	/	/	/	/	/	/	/	/	/	/
G4PromFinder	/	/	/	/	/	/	/	/	/	/	/
G4RNA screener	/	/	/	/	/	/	/	/	/	/	/
G4P Calculator	/	/	/	/	/	/	/	/	/	/	/
G4-iM Grinder	/	/	/	/	/	/	/	/	/	/	/
G4Hunter	/	/	/	/	/	/	/	/	/	/	/
C4P Predictor V2	/	/	/	/	/	/	/	/	/	/	/

method. After connecting the runs for higher-order structures, it calculates their frequencies in the input sequence. After the analysis, G4-iM Grinder counts predefined patterns provided by the users (both single and multiple nucleobase patterns are accepted) in all found PQS. Program can also compare found PQS with validated in vitro G4 structures. Evaluation of the results is prepared based on specified scoring methods, G4Hunter, PQS-finder (incorporated and modified using ML method) and/or cGcC. The final score is computed as a weighted average of the selected scoring systems. Input data format: FASTA. Availability: R package (<https://github.com/EfresBR/G4iMGrinder>).

G4P Calculator [56] finds PQS based on the density of guanine runs in a nucleic acid sequence. Although designed for DNA, it accepts also RNA sequences. The algorithm moves window frames along the sequence and counts frames that meet the specified criteria. Users can define window size, window shift, the minimum length of the G-run and the minimum number of G-runs per window. Default settings: window size = 100 nts, window shift = 20 nts, G-run  $\geq$  3 and G-runs per window  $\geq$  4. Input data format: FASTA. Availability: standalone software (<http://depts.washington.edu/maizels9/G4calc.php>).

G4Predict searches for intramolecular and intermolecular G4s based on a sequence motif. It extends the functionality of Quadparser [52]. The pattern to search for intramolecular G4s is defined as  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$ . Users can determine if the overlapping sequences should be preserved or merged, they can define scores for the number of tetrads, loop lengths and the number of bulges. The bulge score factor is only available for the intramolecular G4 search. Users can also limit the loop size and guanines in the loops. In the intramolecular mode, users can determine the number of bulges ( $\leq 1$  per tetrad) and bulge length. In the intermolecular mode, users can limit the G-runs used to predict partial G4s. Input data format: FASTA. Availability: Python script (<https://github.com/mparker2/g4predict>).

G4-Predictor V.2 [39] locates non-overlapping G4s on sense and anti-sense strands of a given DNA or RNA sequence. Users can set the maximum length of G4 sequence (10–45), the minimum length of G-tract (2–6) and loop size (0–36). G4-Predictor V.2 is accessible from the Mishra group website along with G4IPDB, the G-quadruplex DNA/RNA Interacting Protein Database. Input data formats: raw sequence, FASTA. Availability: web application (<http://bsbe.iiti.ac.in-bsbe/ipdb/pattern2.php>).

G4PromFinder [57] identifies potential transcription promoters in bacterial genomes. It searches for promoters based on G4 DNA motifs and AT-richness. Designed for DNA sequences, the program can be easily modified to process RNA data. The identification of AT-rich fragments follows scanning for PQS in 50 bp upstream region from the 5' end of found AT-rich elements. Searched pattern is  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$ , where  $x \in [2, 4]$ ,  $y \in [1, 10]$ . The maximum length of the G4 sequence equals 30 nucleotides. Input data format: FASTA. Availability: Python script (<https://github.com/MarcoDiSalvo90/G4PromFinder>).

G4RNA screener [58, 59] aims to predict G4s in RNA sequences. Its ML algorithm is trained on experimentally validated G4s from sequences deposited in the G4RNA database [36]. The webserver version of the program allows submitting data of either 20,000 characters or 30 KB. Users can customize the output data, the program allows displaying a variety of optional features, for example, Ensembl gene ID, G4 strand and start and end position of G4. G4RNA screener incorporates consecutive G/C ratio threshold (cGcC), G4Hunter threshold (G4H) and G4 Neural Network threshold (G4NN). In the webserver version, these thresholds are set by default to 4.5, 0.9 and 0.5, respectively. The result table can be downloaded as XLSX or CSV file. Input

data format: FASTA. Availability: Python script, web application ([http://scottgroup.med.usherbrooke.ca/G4RNA\\_screener](http://scottgroup.med.usherbrooke.ca/G4RNA_screener)).

ImgQfinder [60, 61] detects canonical and non-canonical PQS. Depending on users' preferences, it finds either guanine-based or cytosine-based quadruplexes. The search criteria rely on G-runs customized by users into a sequence pattern. Users can set the following parameters: the number of tetrads (2–10), the maximum loop length (2–25), canonical/non-canonical structure (0–1, where 1 implies one bulge or mismatch in G-run) and displaying overlapping/non-overlapping PQS. Non-canonical G4s that contain mismatch in G-run are represented by the search pattern  $G_{i-1}NG_{n-i}$ , bulged G-runs are defined as  $G_{i-1}NG_{n-i+1}$ , where  $i \in (1, n)$ ,  $n \geq 3$ .  $n$  denotes the number of Gs in a G-run;  $i$  shows the position of bulge or mismatch in a G-run. Input data formats: raw sequence, FASTA. Availability: web application (<http://imgqfinder.niifhm.ru>). pqsfinder [35] searches for PQS in DNA or RNA sequences using regular expression for G-runs with length limitation. The searched motif defines as  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$ , where  $x \in [1, 10]$ ,  $y \in [0, 9]$ . Mismatches, bulges and long loops within the searched G-run motif are allowed. pqsfinder offers options within three categories: filters, scoring systems and advanced options. Filters allow selecting strands where PQS are searched; searching for overlapping G4s; limiting G-run length; and setting the minimum PQS score, maximum PQS length, loop size, the maximum number of bulges, mismatches and overall defects. The scoring scheme is based on the stability of the potential G4 structure. G-runs are evaluated individually, with bonus points for each G-tetrad stacking, and penalty points for each mismatch or bulge that occurs in the motif. In the scoring system, the users can set penalties and bonus points for complete G-tetrads. A regular expression, custom scoring and default scoring system can be set by users as advanced options. Input data format: FASTA. Availability: R package (Bioconductor) (<https://bioconductor.org/packages/release/bioc/html/pqsfinder.html>).

QGRS Mapper [62] finds putative G-quadruplexes within DNA or RNA sequences using a sequence motif. QGRS Mapper accepts A, C, T, G, U and N in the input sequence. It searches for pattern  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$ ,  $x \geq 2$ , fixed for all G-tracts. The users can set the maximum length of the quadruplex sequence (10–45), the minimum number of tetrads (2–6), loop size (0–36) and loop sequence, which makes program find at least one loop matching the defined character string. Loop sequence is provided as a regular expression, for example,  $a4$ , (a loop with 4 or more consecutive adenines). The sequence is scored and cleared of the overlapping PQS. The scoring function favors shorter, regular (equal in size) loops over the longer, irregular ones; and relies on the assumption that more stable quadruplexes have more G-tetrads. The results of sequence analysis are displayed as sequence view, data view, data view with overlapping quadruplexes and graphics view. The last one requires the Java Plugin installed on the local machine. The results can be downloaded as an Excel file. Input data formats: raw sequence, FASTA. Availability: web application (<http://bioinformatics.ramapo.edu/QGRS/analyze.php>).

QPARSE [42] is a graph-based algorithm to search for non-canonical G-quadruplexes. It constructs and then traverses a direct acyclic graph of discovered runs. QPARSE finds multimeric potential quadruplex-forming sequences, long-looped PQS and intramolecular monomeric quadruplexes. It applies the mfold-derived function to score the predicted loops based on their thermodynamic and conformation stability. The users can specify searched base in a run (default: G), run length, the maximum loop distance between runs within the same PQS, the number of consecutive runs in the same PQS, the minimum number of

uninterrupted runs and the maximum number of long loops ( $\geq 7$  nt) per each PQS. Either one loop symmetry—mirror or palindrome—or both can be verified within input data. QPARSE webserver limits the input sequence up to 10,000 nucleotides or 15 KB of data. Input data format: FASTA. Availability: Python script, web application (<https://github.com/B3rse/qparse>).

Quadron [37] applies an ML model to predict PQS in DNA sequences. It was developed based on a tree gradient boosting machine. Human genome G4-seq sequences were divided into two sets: one was used as a training set (70% random sequences), and the other one served as a testing set (remaining 30% of the sequences). The general motif in Quadron search defines as  $G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$ . Despite the model was dedicated to DNA sequences, it also handles RNAs. The users specify how many CPUs should algorithm use for calculation. Input data format: FASTA. Availability: R package (<http://quadron.atgcdynamics.org>). The program requires an installation of an R and xgboost library. For users convenience, GUI is also available.

TetraplexFinder [50] searches for potential G-quadruplexes within DNA or RNA sequences. It is a partial module of QuadBase2 [49]—a webserver for PQS prediction within eukaryotic and prokaryotic sequences and user-provided sequences. The tool accepts 20 MB of data, which allows processing large datasets, differentiating whether the input file has a single or multiple entries. The users can set up G-tract length (2–5), loop size (1–50), strand where algorithms should search for PQS, bulge size (0–7) and the algorithm to be used (greedy/non-greedy). Bulges are searched only in GGG tetrads. The output data can be filtered on the website or downloaded as a BED file to the local machine. Input data format: FASTA. Availability: web application (<http://quadbase.igib.res.in/TetraPlexFinder>).

### Tools that predict 2D structure with G4s

In the vast collection of programs that predict RNA secondary structure, only one refers to quadruplexes. RNAfold [63, 64] predicts the secondary structure of RNA or DNA and annotates potential quadruplexes in it. This core program of the ViennaRNA Package applies the thermodynamics-based function to optimize the structure. Additional option—Incorporate G-Quadruplex formation into the structure prediction algorithm—turned on makes the algorithm search for quadruplexes during the computational process. RNAfold outputs the secondary structure in a dot-bracket notation with ‘+’ signs under guanines predicted to form the G-quadruplex. Input data formats: raw sequence, FASTA. Availability: standalone program, web application (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

### Tools that analyze and visualize 2D and 3D structure

In 2017, the first computational tools appeared that could visualize the topology of quadruplexes and reveal their specificity based on the 3D structure data [65, 66]. Shortly afterward, bioinformatics developed methods to identify quadruplexes in nucleic acid structures and determine selected parameters characteristic of these motifs. Currently, four tools can analyze and visualize G4 structures. Their important features are summarized in Table 3: strand polarity indicates whether the tool outputs the information about strands' directions; G4 classification informs what is the basis for classifying the quadruplex topology; base-pair classification indicates if base pairs are classified according to known nomenclatures; area tells if the program calculates surface area of the tetrads; rise

and twist denote that the program computes these parameters for each pair of neighboring tetrads in the quadruplex; planarity checked means that the program analyzes planarity deviation for every tetrad; torsion angles tell that the program outputs torsion angles for the structure; 2D view and 3D view indicate whether the tool visualizes the secondary and tertiary structure; moving camera means that the program allows rotating the visualized 3D model.

DSSR [38] processes the 3D structure of the RNA molecule and annotates its secondary structure. It is a part of the 3DNA suite [67] designed to work with the structures of nucleic acids. DSSR identifies, classifies and describes base pairs, multiplets and characteristic motifs of the secondary structure; helices, stems, hairpin loops, bulges, internal loops, junctions and others. It can also detect modules and tertiary structure patterns, including pseudoknots and kink-turns. The recent extension, DSSR-PyMOL [68], allows drawing cartoon-block schemes of the 3D structure and responds to the need for simplified visualization of quadruplexes. Input data formats: PDB, mmCIF and PDB ID. Availability: standalone program, web application (<http://dssr.x3dna.org>, <http://skmatic.x3dna.org>).

ElTetrad [31] specializes in identifying and describing tetrads and quadruplexes in the 3D structures of nucleic acids, by searching for G-based and non-G-based motifs. It classifies tetrads and quadruplexes into ONZ classes according to their secondary structure topology [2] and calculates strand direction, planarity deviation, rise and twist parameters. The program also outputs the graphical representation of the secondary structure (top-down arc diagram) and its dot-bracket encoding in a two-line format—both designed specially to handle quadruplexes. Input data formats: PDB, mmCIF. Availability: Python script (<https://github.com/tzok/eltetrad>).

RNApDbee [66, 69], a multifunctional tool from RNAPolis suite [70], mainly aims to annotate secondary structures of knotted and unknotted RNAs based on the 3D structure data. Its usefulness in the study of quadruplexes lies in the appropriately matched visualization of the secondary structure, which facilitates the visual identification of these motifs on a diagram of the entire structure and highlights their topological features. A pictographic annotation of interactions within tetrads—according to Leontis-Westhof nomenclature—allows the immediate determination of the tetrad (and quadruplex) class in the recently developed ONZ taxonomy [2]. Input data formats: PDB, mmCIF, PDB ID, BPSEQ, CT, dot-bracket. Availability: web application (<http://rnapdbee.cs.put.poznan.pl/>).

3D-NuS [65] models and visualizes the 3D nucleic acid structures, including duplexes, triplexes and quadruplexes. It builds energy minimized 3D models of canonical and non-canonical G4 structures based on 17 classes defined for the intramolecular and intermolecular quadruplexes. The users provide strand orientation and type, the number of G-quartets, sequences of all G4-loops, to get the model visualized in JSmol along with selected structure data. Input data formats: G4 class, strand type, number of G-tetrads, loops' sequences. Availability: web application (<http://iith.ac.in/3dnus/Quadruplex.html>).

## Results

### Test sets

We created four sets of RNA sequences with and without the ability to form quadruplexes (Figure 2). We used them to test sequence-based tools for the prediction of G4 location in the sequence and prediction of RNA secondary structure

**Table 3.** Selected features of 2D and 3D structure analyzing tools

Tool	Strand polarity	G4 classification	Base-pair classification	Area	Rise	Twist	Planarity	Torsion angles	2D view	3D view	Moving camera
DSR	✓	Loop-based	Saenger, Leontis-Westhof	✓	✓	✓	✓	✓	✓	✓	✓
E1Tetrado	✓	ONZ	Leontis-Westhof	✓	✓	✓	✓	✓	✓	✓	✓
RNApdbee			Saenger, Leontis-Westhof	✓	✓	✓	✓	✓	✓	✓	✓
3D-NuS		Q1-Q17									

with quadruplexes. To build the first test set (DP: dataset with positive examples), we searched the G4RNA database [36] for the sequences, for which the experiments confirmed that they formed G-quadruplexes. We found 321 examples, and after removing the redundant data (unnecessary duplicates), we obtained DP with 295 positive cases. The duplicates were identified and removed by using the MS Excel function *remove duplicates*. In the same database, we found 238 RNAs that did not tend to fold into quadruplexes. By selecting unique sequences, we ended up with 237 negative cases in the DN set (DN: dataset with negative examples). G4RNA is the only database of sequences experimentally tested for G-quadruplex folding, which contains both: G4 sequences and sequences confirmed not to form G4s. Therefore, we chose it as the test data source. Two other test sets were built based on miRBase [71]—the database collecting annotated pre-miRNA and mature miRNA sequences of various species; currently, about 40,000 pre-miRNAs from 271 organisms. We created DH (DH: dataset with human pre-miRNAs) containing 1864 non-redundant sequences selected from 1917 human pre-miRNAs and DV (DV: dataset with *Viridiplantae* pre-miRNAs) with 8354 unique sequences selected from 8615 *Viridiplantae* pre-miRNAs. DH and DV sets contain sequences with quadruplex forming propensity, although their formation was not confirmed experimentally. The data to create all test sets were collected from both repositories on 21 March 2020. More information on the datasets is available in the Supplementary Material (Table S1).

To test the tools that analyze and visualize 2D and 3D structures, we selected two PDB-deposited RNA structures that formed G-quadruplexes: an RNA aptamer (PDB id: 2RQJ) [72] with canonical G4 topology and r(GGAGGAGGAGGA) sequence and a synthetic construct of r(UGGUGGU)4 structure (PDB id: 6GE1) [3] containing U-tetrads.

### Computational experiments with sequence-based tools

In this part of our study, we processed four test sets with 14 tools aimed to predict G4 locations in RNA sequences and one for RNA secondary structure prediction with G4s. Most programs performed with the default settings, apart from G4Catchall tested in two modes, G4RNA screener in four and TetraplexFinder in three. 2 tetrads and 3 tetrads modes of G4Catchall mean that the tool searches for two- or three-tetrad motifs, respectively. G4Predict in the *Intra* mode looks for intramolecular G-quadruplexes. cGcC, G4H and G4NN are thresholds used in G4RNA screener; all mode aggregates the results obtained for all of them. G2 L1-12, G3 L1-7 and G3 L1-3 modes of TetraplexFinder correspond to three stringency levels in PQS search: low (two-tetrad G4s, loop size 1-12), medium (three-tetrad G4s, loop size 1-7) and high (three-tetrad G4s, loop size 1-3). G4-iM Grinder was launched with parameters: Name = LmajorESTs, Sequence = Sequence, DNA = FALSE, RunComposition=G, MinRunSize = 1, MinNRuns = 1, MinPQSSize = 1, Complementary = FALSE. RNAfold was executed with the advanced option *Incorporate G-Quadruplex formation into the structure prediction algorithm*, which set it to the G4 search mode. We post-processed the results using in-house scripts [73, 74] in Python, bash and R.

We executed each tool for each sequence in the DP, DN, DH and DV sets, and we noted whether the tool predicted quadruplexes in it or not. Within every test set, we calculated the number of G4-positive and G4-negative sequences found by each tool (Supplementary Material, Tables S2 and S3, Figure S1), where a sequence with predicted quadruplex is counted as G4-positive, and a sequence without quadruplex is G4-negative.

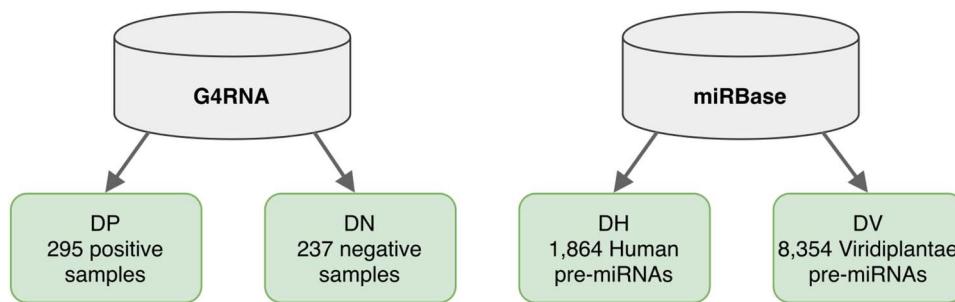


Figure 2. Test sets created for the analysis of sequence-based tools.

Table 4. Test set coverage [%] with PQS predictions: positive (+) for all sets, negative (-) for DP and DN

Tool	Mode	DP+	DP-	DN+	DN-	DH+	DV+
G4Catchall	2 tetrads	91.5	8.5	40.5	59.5	15.7	4.2
	3 tetrads	86.4	13.6	30.8	69.2	9.0	1.4
G4Hunter	n/a	97.6	2.4	67.5	32.5	53.9	44.4
G4-iM Grinder	n/a	100.0	0.0	97.5	2.5	93.2	92.1
G4P Calculator	n/a	78.0	22.0	38.8	61.2	10.2	3.2
G4Predict	Intra	14.6	85.4	2.5	97.5	0.6	0.1
G4-Predictor V.2	n/a	99.0	1.0	66.7	33.3	44.2	30.9
G4PromFinder	n/a	28.1	71.9	30.8	69.2	14.8	19.6
G4RNA screener	cGcC	85.1	14.9	40.1	59.9	8.3	21.0
	G4H	81.0	19.0	25.7	74.3	3.1	1.1
	G4NN	84.4	15.6	30.0	70.0	6.0	3.3
	all	95.6	4.4	46.8	53.2	10.9	22.2
ImGQfinder	n/a	16.3	83.7	5.5	94.5	0.8	0.1
pqsfinder	n/a	86.8	13.2	36.7	63.3	10.6	1.4
QGRS Mapper	n/a	99.0	1.0	66.2	33.8	44.2	30.9
QPARSE	n/a	97.3	2.7	65.0	35.0	43.0	29.8
Quadron	n/a	70.5	29.5	21.1	78.9	5.8	0.9
TetraplexFinder	G2 L1-12	39.3	60.7	29.1	70.9	5.4	1.7
	G3 L1-7	14.6	85.4	2.1	97.9	0.4	0.0
	G3 L1-3	9.2	90.8	1.3	98.7	0.4	0.0
RNAfold	Advanced	73.2	26.8	21.9	78.1	2.5	0.8

In Table 4, we show the coverage of all sets with positive (+) predictions and, additionally, the coverage of DP and DN with the negative (-) ones. Let us recall that the DP set contains sequences for which experiments confirmed the formation of quadruplexes, sequences from DN do not form quadruplexes, DH and DV include sequences with quadruplex forming propensity. We expected the best tools to show high coverage of DP with positive PQS predictions and DN with negative predictions. As shown in Figure 3, several programs meet these conditions. The best results achieved G4RNA screener, G4Catchall and RNAfold. G4RNA screener (in all modes) identified PQS in >80% sequences in DP and classified >50% of DN sequences as non-PQS; however, note that this algorithm was trained on data from G4RNA database (as of 2017) and this result was expected. G4Catchall generated around 90% of positive predictions for DP and 60–70% negative ones for DN. RNAfold showed over 70% of correct predictions for both sets. Just behind these three programs are Quadron and pqsfinder—both cover >60% of DP and DN with correct predictions. Relatively few PQS were found in the DV and DH datasets, which contain sequences potentially forming quadruplexes. In the vast majority of cases, the coverage of these sets with positive predictions does not exceed 10%.

A separate group of tools maximize the number of predicted PQS. Among them, G4-iM Grinder stands out in the foreground—it found quadruplexes in all sequences of DP and 97.5% sequences of DN. The opposite strategy is adopted by G4Predict, ImGQfinder and TetraplexFinder, which in all data sets found few sequences with the potential to create quadruplexes. In most cases, these programs give at most 15% coverage with correct predictions.

Finally, G4PromFinder surprisingly recognizes more PQS in DN than DP set. This program addresses large sequences (bacterial genomes) where it searches for potential promoters. Therefore, the input sequence length should exceed 50 nucleotides, with >40% of adenines and uracils, and motif length  $\leq$  30 nucleotides. In the DP test set, 147 of the 295 sequences (49.8%) consist of  $\geq$  50 nucleotides, while in DN, 177 of the 237 (74.7%). G4PromFinder predicted 83 PQS in DP set (28.1%) and 73 PQS in the DN set (30.8%). Such predictions are the result of the program's prerequisites.

Based on the results obtained for DP and DN sets, we evaluated the quality of prediction (Table 5) by computing the following:

- accuracy:  $ACC = \frac{TP+TN}{TP+TN+FP+FN}$
- sensitivity (true positive rate):  $TPR = \frac{TP}{TP+FN}$



Figure 3. Coverage of DP and DN datasets with correct predictions: positive in DP and negative in DN [%].

- specificity (true negative rate):  $TNR = \frac{TN}{TN+FP}$
- precision (positive predictive value):  $PPV = \frac{TP}{TP+FP}$
- negative predictive value:  $NPV = \frac{TN}{TN+FN}$
- false discovery rate:  $FDR = \frac{FP}{TP+FP}$
- F-score:  $F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR}$

They all use four basic measures: true positives (TP)—PQS predicted for DP sequences, true negatives (TN)—negative predictions in DN set, false positives (FP)—PQS predicted for DN sequences and false negatives (FN)—negative predictions in DP set (Supplementary Material, Table S4).

Accuracy (ACC) is the ratio of correct predictions to the total number of input sequences. G4Catchall, G4RNA screener and pqsfinder have the best (the highest) accuracy, G4PromFinder—the worst (the lowest) one, which confirms our observations from the previous paragraphs. Sensitivity (TPR) indicates what part of the actual PQS has been predicted by the program. The highest TPR (i.e. the best one) belongs to G4-iM Grinder, G4-Predictor V.2 and QGRS Mapper, but we already know that these tools aim to maximize the prediction of PQS. Such a strategy also causes poor (the lowest) specificity (TNR). Low TNR value indicates a small fraction of correctly predicted PQS-negative sequences. TetraplexFinder, ImGQfinder and G4Predict are the leaders of specificity, with its value exceeding 0.95. TetraplexFinder and G4Predict have also the highest precision (PPV). PPV shows a fraction of positive predictions, which are natively positive. In turn, NPV determines which part of the negative predictions is actually negative. A reliable tool takes high values of all mentioned factors, but especially PPV and NPV should be close to 1. False discovery rate (FDR) is the only measure from Table 5 to be minimized. It points the fraction of incorrectly predicted PQS among all positive predictions. The lowest FDR belongs to TetraplexFinder and G4Predict. Finally, the F-score, a weighted harmonic mean of precision and sensitivity, is computed to find the balance between these two measures.

It aims to assess the accuracy when the distribution between classes is uneven—especially if there is a large number of true negatives. The best (highest) F-score has G4Catchall and G4RNA screener, with pqsfinder right behind, while the worst F-score belongs to TetraplexFinder and G4Predict—two tools showing the best precision. In Table 5, the best value of each computed measure is highlighted in bold.

#### Computational experiments with structure-based tools

In this part of the study, we tested four tools that analyze and visualize the secondary and the tertiary structures with quadruplexes. We executed them for two quadruplex RNA structures—an RNA aptamer (PDB id: 2RQJ) [72] and a synthetic construct of r(UGGUGGU)4 structure (PDB id: 6GE1) [3]—to find out what details of the structure we can obtain and in what form, numerical and graphical, they are presented. Let us note that 2RQJ is a dimer with two unimolecular structures obtained via NMR; each comprising one G-quadruplex with two tetrads of O+ type in the ONZ classification [2]. 6GE1 is an NMR-determined, tetramolecular structure with unusual topology. It is composed of seven tetrads—four G-tetrads and three U-tetrads—six of them classified as O+ and one as O- according to the ONZ taxonomy. ElTetraido and RNAPdbee were run with the default input settings, DSSR with the PDB identifier at the input only, 3D-NuS required to select the quadruplex class.

DSSR and ElTetraido identified quadruplexes in the input PDB files. Both programs focused on structural aspects of the input molecule, explicitly informing about quadruplexes and tetrads within the structure. DSSR provided an extensive analysis of 3D structures and output the data about G-tetrads, G-helices and G4-stems. It computed planarity for each G-tetrad and gave the sections area, rise and twist parameters for G4-helix and G4-stems. The program automatically assigned loop topologies

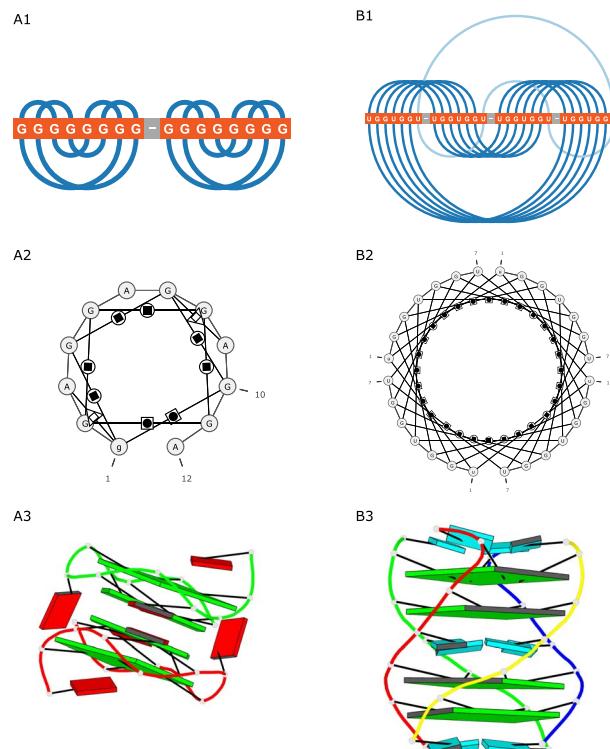
**Table 5.** Quality of PQS prediction based on DP and DN set processing

Tool	Mode	ACC	TPR	TNR	PPV	NPV	FDR	F1
G4Catchall	G2	0.77	0.92	0.59	0.74	0.85	0.26	<b>0.82</b>
	G3	<b>0.79</b>	0.86	0.69	0.78	0.80	0.22	<b>0.82</b>
G4Hunter	n/a	0.69	0.98	0.32	0.64	0.92	0.36	0.78
G4-iM Grinder	n/a	0.57	<b>1.00</b>	0.03	0.56	<b>1.00</b>	0.44	0.72
G4P Calculator	n/a	0.70	0.78	0.61	0.71	0.69	0.29	0.75
G4Predict	Intra	0.52	0.15	0.97	0.88	0.48	0.12	0.25
G4-Predictor V.2	n/a	0.70	0.99	0.33	0.65	0.96	0.35	0.78
G4PromFinder	n/a	0.46	0.28	0.69	0.53	0.44	0.47	0.37
G4RNA screener	cGcC	0.74	0.85	0.60	0.73	0.76	0.27	0.78
	G4H	0.78	0.81	0.74	0.80	0.76	0.20	0.80
	G4NN	0.78	0.84	0.70	0.78	0.78	0.22	0.81
	all	0.77	0.96	0.53	0.72	0.91	0.28	<b>0.82</b>
ImGQfinder	n/a	0.51	0.16	0.95	0.79	0.48	0.21	0.27
pqsfinder	n/a	0.76	0.87	0.63	0.75	0.79	0.25	0.80
QGRSMapper	n/a	0.70	0.99	0.34	0.65	0.96	0.35	0.78
QPARSE	n/a	0.70	0.97	0.35	0.65	0.91	0.35	0.78
Quadron	n/a	0.74	0.71	0.79	0.81	0.68	0.19	0.75
TetraplexFinder	G2 L1-12	0.53	0.39	0.71	0.63	0.48	0.37	0.48
	G3 L1-7	0.52	0.15	0.98	<b>0.90</b>	0.48	<b>0.10</b>	0.25
	G3 L1-3	0.49	0.09	<b>0.99</b>	<b>0.90</b>	0.47	<b>0.10</b>	0.17
RNAfold	Advanced	0.75	0.73	0.78	0.81	0.70	0.19	0.77

according to the predefined types (P—parallel, D—diagonal and L—lateral) and their orientation (+/−). DSSR-PyMOL generated block schemes of both quadruplexes (**Figure 4A3** and **B3**). ElTetra also calculated planarity, rise and twist parameters and identified strand directions for both quadruplexes. It classified the quadruplexes and their component tetrads to ONZ classes. Finally, it generated the arc diagram (**Figure 4A1** and **B1**) and two-line dot-bracket encoding of every quadruplex.

RNApdbee, as opposed to the previous programs, does not explicitly inform that it has identified tetrads and quadruplexes in the input data. Its purpose is to annotate and visualize the secondary structure and determine its parameters, focusing on pseudoknots [66, 69]. For the analyzed structures, RNApdbee generated an extensive report on the secondary structure, including information on canonical and non-canonical interactions, their classification in Saenger and Leontis–Westhof nomenclatures, base-phosphate interactions, stacking interactions, base-ribose interactions, structure motifs, dot-bracket and the secondary structure diagram—drawn by VARNA-based procedure—with base pairs annotated according to Leontis–Westhof [75] (**Figure 4A2** and **B2**). Note that only one, VARNA-based, drawing procedure of RNApdbee can visualize quadruplexes. The other two, PseudoViewer-based procedure and R-chie, are based on canonical interactions and their visualizations of quadruplex structures are incomplete.

3D-NuS aims to generate the 3D models of G4s based on 17 classes of G-quadruplex folds. Thus, its input and output data differ from the other tools in this section. The program requires input information about quadruplex topology: quadruplex class, subclass and the number of tetrads and sequences of loops. It outputs the tertiary structure in PDB format and provides its visualization. We tried 3D-NuS for different quadruplexes, including the ones selected for the analysis. Tests have shown that 3D-NuS is not a suitable tool for modeling RNA quadruplexes. It does not form non-G tetrads, it has problems with modeling short loops, uni- and bimolecular quadruplexes; no such observations appeared when it modeled DNA quadruplexes with similar sequence and topology. Provided input data and



**Figure 4.** Visualization of (A) 2RQJ and (B) 6GE1 structures generated by (1) ElTetra, (2) RNApdbee and (3) DSSR-PyMOL.

output data generated by 3D-NuS for exemplary structures are presented in the [Supplementary Material](#).

## Conclusion

With the growing interest in quadruplexes, computer programs for their analysis began to appear. Most of them rely solely

on a sequence and parse it to find a predefined G4 motif. This goes hand in hand with creating G4-related databases that primarily collect information about sequences with the ability to form quadruplexes. Our experiments with sequence-based tools applied for RNA sequences showed a very good performance of G4Catchall (motif-based algorithm), whose flexibility certainly contributed to this result. Right behind was RNAfold, the tool for secondary structure prediction enriched with the quadruplex annotation option. Four existing structure-based tools addressing G4s focus on different structural aspects. DSSR comprehensively examines the G4 structure, determines a variety of its parameters and provides the schematic 3D view. ElTetrad identifies tetrads and quadruplexes in the structure, computes their basic parameters, classifies according to ONZ taxonomy and gives the secondary structure in the arc diagram and dot-bracket notation. RNAPdbee draws secondary structure diagrams and classifies base-pairs. 3D-NuS builds the 3D model of the quadruplex based on user-defined topology if the quadruplex topology fits one of the classes supported by the tool. These tools complement each other in revealing the full picture of quadruplex space, although they do not deal equally well with all quadruplex types, e.g. 3D-NuS is limited to 17 G4 classes and can reliably model DNA quadruplexes only.

Despite the already significant number of bioinformatics programs that can be used to study DNA and RNA quadruplexes, there are still issues that lack *in silico* solutions. One of them is the modeling of the secondary structure. Among the huge number of programs to predict the RNA 2D structure, only RNAfold touched the problem of quadruplexes. It annotates the places of G4 formation in the dot-bracket representation of the structure. However, this notation does not reflect the quadruplex topology and cannot be easily transformed into secondary structure visualization. Prediction and modeling of the quadruplex 3D structure is also a challenge. First reported attempts of blind, template-free prediction of the 3D G4 structure were made within RNA-Puzzles challenge 23, with seven human and one webserver participant. From the assessment table (available online), it can be seen that this structure was one of the most difficult ones in RNA-Puzzles history. A reliable prediction of the 2D and 3D structure of quadruplexes requires experimental data inclusion, e.g. thermodynamic parameters. A resource collecting such data for G4s could be very supportive. A database that would integrate various data from existing archives would also be a helpful tool or a specialized search engine, browsing the existing databases for related information on a given quadruplex.

### Key points

- G4-related computational tools concentrate on discovering, analyzing, and visualizing quadruplexes, mostly on the sequence level.
- In this work, we analyzed 35 bioinformatics resources: 10 dedicated solely to DNA G4s; 4 for RNA G4s; 21 for any nucleic acid.
- Tests of existing G4-related sequence-based tools against four RNA datasets identified G4Catchall, a motif-based method, as the best tool for finding reliable RNA PQS.
- Only 3 tools analyze the 2D and 3D structures of RNA quadruplexes. Their functions are complementary: each considers the other set of structural features and generates a different view of the G4 topology.

- The sequence-based modeling of the G4 structure is challenging. Only one program for 2D structure prediction—RNAfold—reliably indicates G4 location in the sequence but does not give its topology. The 3D structure prediction of RNA G4s is still a challenge.

### Funding

National Science Centre, Poland (2016/23/B/ST6/03931, 2019/35/B/ST6/03074); Poznań University of Technology.

### References

1. Lightfoot HL, Hagen T, Tatum NJ, et al. The diverse structural landscape of quadruplexes. *FEBS Lett* 2019;593(16):2083–102. doi: [10.1002/1873-3468.13547](https://doi.org/10.1002/1873-3468.13547).
2. Popenda M, Miskiewicz J, Sarzynska J, et al. Topology-based classification of tetrads and quadruplex structures. *Bioinformatics* 2020;36(4):1129–34. doi: [10.1093/bioinformatics/btz738](https://doi.org/10.1093/bioinformatics/btz738).
3. Andrałojć W, Małgowska M, Sarzyńska J, et al. Unraveling the structural basis for the exceptional stability of RNA G-quadruplexes capped by a uridine tetrad at the 3' terminus. *RNA* 2018;25(1):121–34. doi: [10.1261/rna.068163.118](https://doi.org/10.1261/rna.068163.118).
4. Zhang N, Gorin A, Majumdar A, et al. Dimeric DNA quadruplex containing major groove-aligned A-T-A-T and G-C-G-C tetrads stabilized by inter-subunit Watson-crick a-T and G-C pairs. *J Mol Biol* 2001;312(5):1073–88. doi: [10.1006/jmbi.2001.5002](https://doi.org/10.1006/jmbi.2001.5002).
5. Hedd B, Martín-Pintado N, Serimbetov Z, et al. G-quadruplexes with  $(4n - 1)$  guanines in the G-tetrad core: formation of a G-triad·water complex and implication for small-molecule binding. *Nucleic Acids Res* 2015;44(2):910–6. doi: [10.1093/nar/gkv1357](https://doi.org/10.1093/nar/gkv1357).
6. Kettani A, Gorin A, Majumdar A, et al. A dimeric DNA interface stabilized by stacked A·(G · G · G · G). A hexads and coordinated monovalent cations. *J Mol Biol* 2000;297(3):627–44. doi: [10.1006/jmbi.2000.3524](https://doi.org/10.1006/jmbi.2000.3524).
7. Kogut M, Kleist C, Czub J. Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail. *PLoS Comput Biol* 2019 e1007383; 15(9). doi: [10.1371/journal.pcbi.1007383](https://doi.org/10.1371/journal.pcbi.1007383).
8. Kolesnikova S, Curtis EA. Structure and function of multimeric G-quadruplexes. *Molecules* 2019;24(17):3074. doi: [10.3390/molecules24173074](https://doi.org/10.3390/molecules24173074).
9. Webba da Silva M. Geometric formalism for DNA quadruplex folding. *Chem A Eur J* 2007;13(35):9738–45. doi: [10.1002/chem.200701255](https://doi.org/10.1002/chem.200701255).
10. Lech CJ, Hedd B, Phan AT. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res* 2012;41(3):2034–46. doi: [10.1093/nar/gks1110](https://doi.org/10.1093/nar/gks1110).
11. Webba da Silva M, Trajkovski M, Sannohe Y, et al. Design of a G-quadruplex topology through glycosidic bond angles. *Angewandte Chemie* 2009;121(48):9331–4. doi: [10.1002/ange.200902454](https://doi.org/10.1002/ange.200902454).
12. Dvorkin SA, Karsisiotis AI, da Silva MW. Encoding canonical DNA quadruplex structure. *Sci Adv* 2018;4(8): eaat3007. doi: [10.1126/sciadv.aat3007](https://doi.org/10.1126/sciadv.aat3007).
13. Ravichandran S, Ahn J-H, Kim KK. Unraveling the regulatory G-quadruplex puzzle: lessons from genome and

- transcriptome-wide studies. *Front Genet* 2019;10:1002. doi: [10.3389/fgene.2019.01002](https://doi.org/10.3389/fgene.2019.01002).
14. Hänsel-Hertsch R, Spiegel J, Marsico G, et al. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat Protoc* 2018;13(3):551–64. doi: [10.1038/nprot.2017.150](https://doi.org/10.1038/nprot.2017.150).
  15. Chambers VS, Marsico G, Boutell JM, et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 2015;33(8):877–81. doi: [10.1038/nbt.3295](https://doi.org/10.1038/nbt.3295).
  16. Marsico G, Chambers VS, Sahakyan AB, et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res* 2019;47(8):3862–74. doi: [10.1093/nar/gkz179](https://doi.org/10.1093/nar/gkz179).
  17. Raguseo F, Chowdhury S, Minard A, et al. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem Commun* 2020;56(9):1317–24. doi: [10.1039/c9cc09107f](https://doi.org/10.1039/c9cc09107f).
  18. Che T, Wang Y-Q, Huang Z-L, et al. Natural alkaloids and heterocycles as G-quadruplex ligands and potential anticancer agents. *Molecules* 2018;23(2):493. doi: [10.3390/molecules23020493](https://doi.org/10.3390/molecules23020493).
  19. Kwok CK, Marsico G, Sahakyan AB, et al. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* 2016;13(10):841–4. doi: [10.1038/nmeth.3965](https://doi.org/10.1038/nmeth.3965).
  20. Yeung PY, Zhao J, Chow EY-C, et al. Systematic evaluation and optimization of the experimental steps in RNA G-quadruplex structure sequencing. *Sci Rep* 2019;9(1):8091. doi: [10.1038/s41598-019-44541-4](https://doi.org/10.1038/s41598-019-44541-4).
  21. Yang SY, Lejault P, Chevrier S, et al. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat Commun* 2018;9(1):4730. doi: [10.1038/s41467-018-07224-8](https://doi.org/10.1038/s41467-018-07224-8).
  22. Lee DSM, Ghanem LR, Barash Y. Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat Commun* 2020;11(1):527. doi: [10.1038/s41467-020-14404-y](https://doi.org/10.1038/s41467-020-14404-y).
  23. Chan KL, Peng B, Umar MI, et al. Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem Commun* 2018;54(77):10878–81. doi: [10.1039/c8cc04635b](https://doi.org/10.1039/c8cc04635b).
  24. Roxo C, Kotkowiak W, Pasternak A. G-quadruplex-forming aptamers - characteristics, applications, and perspectives. *Molecules* 2019;24(20):3781. doi: [10.3390/molecules24203781](https://doi.org/10.3390/molecules24203781).
  25. Brázda V, Hároníková L, Liao J, et al. DNA and RNA quadruplex-binding proteins. *Int J Mol Sci* 2014;15(10):17493–517. doi: [10.3390/ijms151017493](https://doi.org/10.3390/ijms151017493).
  26. Serikawa T, Spanos C, von Hacht N, et al. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* 2018;144:169–84. doi: [10.1016/j.biochi.2017.11.003](https://doi.org/10.1016/j.biochi.2017.11.003).
  27. Rouleau SG, Garant J-M, Bolduc F, et al. G-quadruplexes influence pri-microRNA processing. *RNA Biol* 2017;15(2):198–206. doi: [10.1080/15476286.2017.1405211](https://doi.org/10.1080/15476286.2017.1405211).
  28. Kang C, Zhang X, Ratliff R, et al. Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature* 1992;356:126–31. doi: [10.1038/356126a0](https://doi.org/10.1038/356126a0).
  29. Cheong C, Moore PB. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* 1992;31:8406–14. doi: [10.1021/bi00151a003](https://doi.org/10.1021/bi00151a003).
  30. Berman HM. The Protein Data Bank. *Nucleic Acids Res* 2000;28(1):235–42. doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
  31. Zok T, Popenda M, Szachniuk M. ElTetrado: a tool for identification and classification of tetrads and quadruplexes. *BMC Bioinformatics* 2020;21:40. doi: [10.1186/s12859-020-3385-1](https://doi.org/10.1186/s12859-020-3385-1).
  32. Joachimi A, Benz A, Hartig JS. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg Med Chem* 2009;17(19):6811–5. doi: [10.1016/j.bmc.2009.08.043](https://doi.org/10.1016/j.bmc.2009.08.043).
  33. Lombardi EP, Londoño-Vallejo A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res* 2020;48(3):1603. doi: [10.1093/nar/gkaa033](https://doi.org/10.1093/nar/gkaa033).
  34. Wong HM, Stegle O, Rodgers S, et al. A toolbox for predicting G-quadruplex formation and stability. *J Nucleic Acids* 2010;2010:1–6. doi: [10.4061/2010/564946](https://doi.org/10.4061/2010/564946).
  35. Hon J, Martínek T, Zendulka J, et al. PQSfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* 2017;33(21):3373–9. doi: [10.1093/bioinformatics/btx413](https://doi.org/10.1093/bioinformatics/btx413).
  36. Garant J-M, Luce MJ, Scott MS, et al. G4RNA: an RNA G-quadruplex database. *Database* 2015. doi: [10.1093/database/bav059](https://doi.org/10.1093/database/bav059).
  37. Sahakyan AB, Chambers VS, Marsico G, et al. Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci Rep* 2017;7(1):14535. doi: [10.1038/s41598-017-14017-4](https://doi.org/10.1038/s41598-017-14017-4).
  38. Lu X-J, Bussemaker HJ, Olson WK. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* 2015;43(21):e142. doi: [10.1093/nar/gkv716](https://doi.org/10.1093/nar/gkv716).
  39. Mishra SK, Tawani A, Mishra A, et al. G4IPDB: a database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci Rep* 2016;6:38144. doi: [10.1038/srep38144](https://doi.org/10.1038/srep38144).
  40. Li Q, Xiang J-F, Yang Q-F, et al. G4db: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res* 2012;41(D1):D1115–23. doi: [10.1093/nar/gks1101](https://doi.org/10.1093/nar/gks1101).
  41. Bedrat A, Lacroix L, Mergny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* 2016;44(4):1746–59. doi: [10.1093/nar/gkw006](https://doi.org/10.1093/nar/gkw006).
  42. Berselli M, Lavezzo E, Toppi S. QPARSE: searching for long-looped or multimeric G-quadruplexes potentially distinctive and druggable. *Bioinformatics* 2019;36(2):393–9. doi: [10.1093/bioinformatics/btz569](https://doi.org/10.1093/bioinformatics/btz569).
  43. Shao X, Zhang W, Umar MI, et al. RNA G-quadruplex structures mediate gene regulation in bacteria. *MBio* 2020;11(1):e02926–19. doi: [10.1128/mbio.02926-19](https://doi.org/10.1128/mbio.02926-19).
  44. Zhang R, Lin Y, Zhang C-T. Greglist: a database listing potential G-quadruplex regulated genes. *Nucleic Acids Res* 2007;36(D1):D372–6. doi: [10.1093/nar/gkm787](https://doi.org/10.1093/nar/gkm787).
  45. Kikin O, Zappala Z, D'Antonio L, et al. GRSDB2 and GRS\_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res* 2007;36(D1):D141–8. doi: [10.1093/nar/gkm982](https://doi.org/10.1093/nar/gkm982).
  46. Lavezzo E, Berselli M, Frasson I, et al. G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLoS Comput Biol* 2018;14(12):e1006675. doi: [10.1371/journal.pcbi.1006675](https://doi.org/10.1371/journal.pcbi.1006675).
  47. Cer RZ, Donohue DE, Mudunuri US, et al. Non-b DB v2.0: a database of predicted non-b DNA-forming motifs and its associated tools. *Nucleic Acids Res* 2012;41(D1):D94–D100. doi: [10.1093/nar/gks955](https://doi.org/10.1093/nar/gks955).
  48. Ge F, Wang Y, Li H, et al. Plant-GQ: an integrative database of G-quadruplex in plant. *J Comput Biol* 2019;26(9):1013–9. doi: [10.1089/cmb.2019.0010](https://doi.org/10.1089/cmb.2019.0010).
  49. Yadav VK, Abraham JK, Mani P, et al. QuadBase: genome-wide database of G4 DNA occurrence and conservation in human, chimpanzee, mouse and rat promoters and

- 146 microbes. *Nucleic Acids Res* 2007;36(D1):D381–5. doi: [10.1093/nar/gkm781](https://doi.org/10.1093/nar/gkm781).
50. Dhapola P, Chowdhury S. QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res* 2016;44(W1):W277–83. doi: [10.1093/nar/gkw425](https://doi.org/10.1093/nar/gkw425).
  51. D'Antonio L, Bagga P. Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences. In: *Proceedings of the IEEE Computational Systems Bioinformatics Conference* 2004. 590–1. Stanford, CA: IEEE. doi: [10.1109/CSB.2004.1332508](https://doi.org/10.1109/CSB.2004.1332508).
  52. Huppert JL. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 2005;33(9):2908–16. doi: [10.1093/nar/gki609](https://doi.org/10.1093/nar/gki609).
  53. Doluca O. G4Catchall: a G-quadruplex prediction approach considering atypical features. *J Theor Biol* 2019;463:92–8. doi: [10.1016/j.jtbi.2018.12.007](https://doi.org/10.1016/j.jtbi.2018.12.007).
  54. Brázda V, Kolomazník J, Lýsek J, et al. G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 2019;35(18):3493–5. doi: [10.1093/bioinformatics/btz087](https://doi.org/10.1093/bioinformatics/btz087).
  55. Belmonte-Reche E, Morales JC. G4-iM grinder: DNA and RNA G-Quadruplex, i-motif and higher order structure search and analyser tool. *NAR Genom Bioinfor* 2020;2(1):lqz005. doi: [10.1093/nargab/lqz005](https://doi.org/10.1093/nargab/lqz005).
  56. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 2006;34(14):3887–96. doi: [10.1093/nar/gkl529](https://doi.org/10.1093/nar/gkl529).
  57. Salvo MD, Pinatel E, Talà A, et al. G4PromFinder: an algorithm for predicting transcription promoters in GC-rich bacterial genomes based on AT-rich elements and G-quadruplex motifs. *BMC Bioinformatics* 2018;19(1):36. doi: [10.1186/s12859-018-2049-x](https://doi.org/10.1186/s12859-018-2049-x).
  58. Garant J-M, Perreault J-P, Scott MS. Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* 2017;33(22):3532–7. doi: [10.1093/bioinformatics/btx498](https://doi.org/10.1093/bioinformatics/btx498).
  59. Garant J-M, Perreault J-P, Scott MS. G4RNA screener web server: user focused interface for RNA G-quadruplex prediction. *Biochimie* 2018;151:115–8. doi: [10.1016/j.biochi.2018.06.002](https://doi.org/10.1016/j.biochi.2018.06.002).
  60. Varizhuk A, Ischenko D, Tsvetkov V, et al. The expanding repertoire of G4 DNA structures. *Biochimie* 2017;135:54–62. doi: [10.1016/j.biochi.2017.01.003](https://doi.org/10.1016/j.biochi.2017.01.003).
  61. Varizhuk A, Ischenko D, Smirnov I, et al. An improved search algorithm to find G-quadruplexes in genome sequences. *bioRxiv* 2014. doi: [10.1101/001990](https://doi.org/10.1101/001990).
  62. Kikin O, D'Antonio L, Bagga PS. QGRS mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* 2006;34(W1):W676–82. doi: [10.1093/nar/gkl253](https://doi.org/10.1093/nar/gkl253).
  63. Gruber AR, Lorenz R, Bernhart SH, et al. The Vienna RNA Websuite. *Nucleic Acids Res* 2008;36(W1):W70–4. doi: [10.1093/nar/gkn188](https://doi.org/10.1093/nar/gkn188).
  64. Lorenz R, Bernhart SH, zu Biederdissen CH, et al. ViennaRNA Package 2.0. *Algorithm Mol Biol* 2011;6(1):26. doi: [10.1186/1748-7188-6-26](https://doi.org/10.1186/1748-7188-6-26).
  65. Patro LPP, Kumar A, Kolimi N, et al. 3d-NuS: a web server for automated modeling and visualization of non-canonical 3-dimensional nucleic acid structures. *J Mol Biol* 2017;429(16):2438–48. doi: [10.1016/j.jmb.2017.06.013](https://doi.org/10.1016/j.jmb.2017.06.013).
  66. Zok T, Antczak M, Zurkowski M, et al. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res* 2018;46(W1):W30–5. doi: [10.1093/nar/gky314](https://doi.org/10.1093/nar/gky314).
  67. Lu X-J, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* 2008;3(7):1213–27. doi: [10.1038/nprot.2008.104](https://doi.org/10.1038/nprot.2008.104).
  68. Lu X-J. DSSR-enabled innovative schematics of 3d nucleic acid structures with PyMOL. *Nucleic Acids Res* 2020;48:e77. doi: [10.1093/nar/gkaa426](https://doi.org/10.1093/nar/gkaa426).
  69. Antczak M, Popenda M, Zok T, et al. New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation. *Bioinformatics* 2018;34(8):1304–12. doi: [10.1093/bioinformatics/btx783](https://doi.org/10.1093/bioinformatics/btx783).
  70. Szachniuk M. RNAPolis: computational platform for RNA structure analysis. *Found Comput Decis Sci* 2019;44(2):241–57. doi: [10.2478/fcds-2019-0012](https://doi.org/10.2478/fcds-2019-0012).
  71. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2018;47(D1):D155–62. doi: [10.1093/nar/gky1141](https://doi.org/10.1093/nar/gky1141).
  72. Mashima T, Matsugami A, Nishikawa F, et al. Unique quadruplex structure and interaction of an RNA aptamer against bovine prion protein. *Nucleic Acids Res* 2009;37(18):6249–58. doi: [10.1093/nar/gkp647](https://doi.org/10.1093/nar/gkp647).
  73. Miskiewicz J, Tomczyk K, Mickiewicz A, et al. Bioinformatics study of structural patterns in plant microRNA precursors. *Biomed Res Int* 2017;6783010. doi: [10.1155/2017/6783010](https://doi.org/10.1155/2017/6783010).
  74. Miskiewicz J, Szachniuk M. Discovering structural motifs in miRNA precursors from Viridiplantae kingdom. *Molecules* 2018;23(6):1367. doi: [10.3390/molecules23061367](https://doi.org/10.3390/molecules23061367).
  75. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA* 2001;7(4):499–512. doi: [10.1017/s1355838201002515](https://doi.org/10.1017/s1355838201002515).