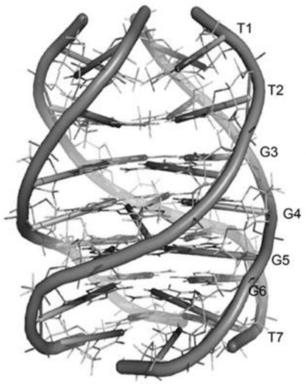
CRAIG WANTS ME TO **TALK ABOUT G4S**



MARTA SZACHNIUK

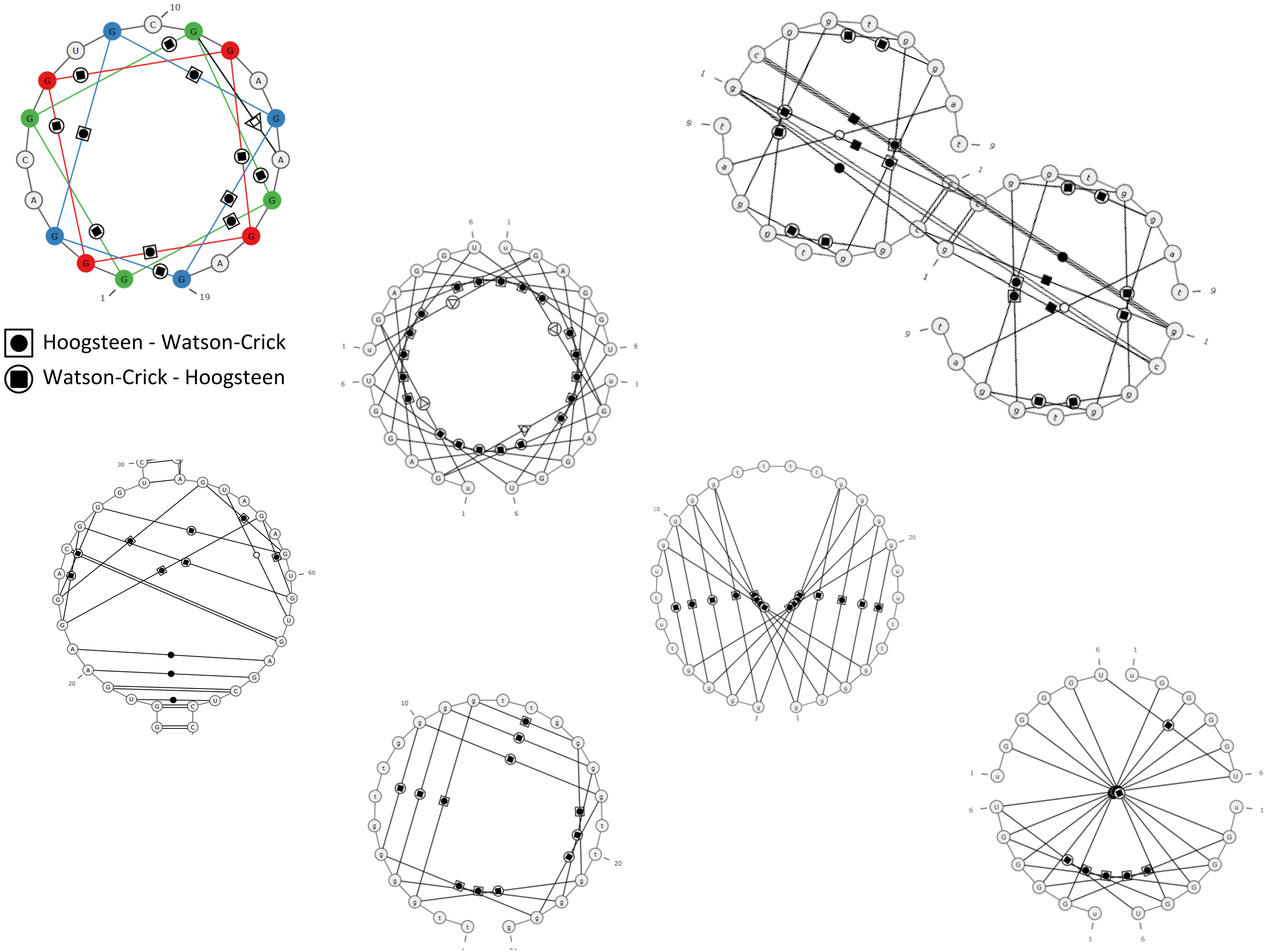
mszachniuk@cs.put.poznan.pl

Institute of Computing Science,
Poznan University of Technology,
Poland

Institute of Bioorganic Chemistry
Polish Academy of Sciences,
Poznan, Poland

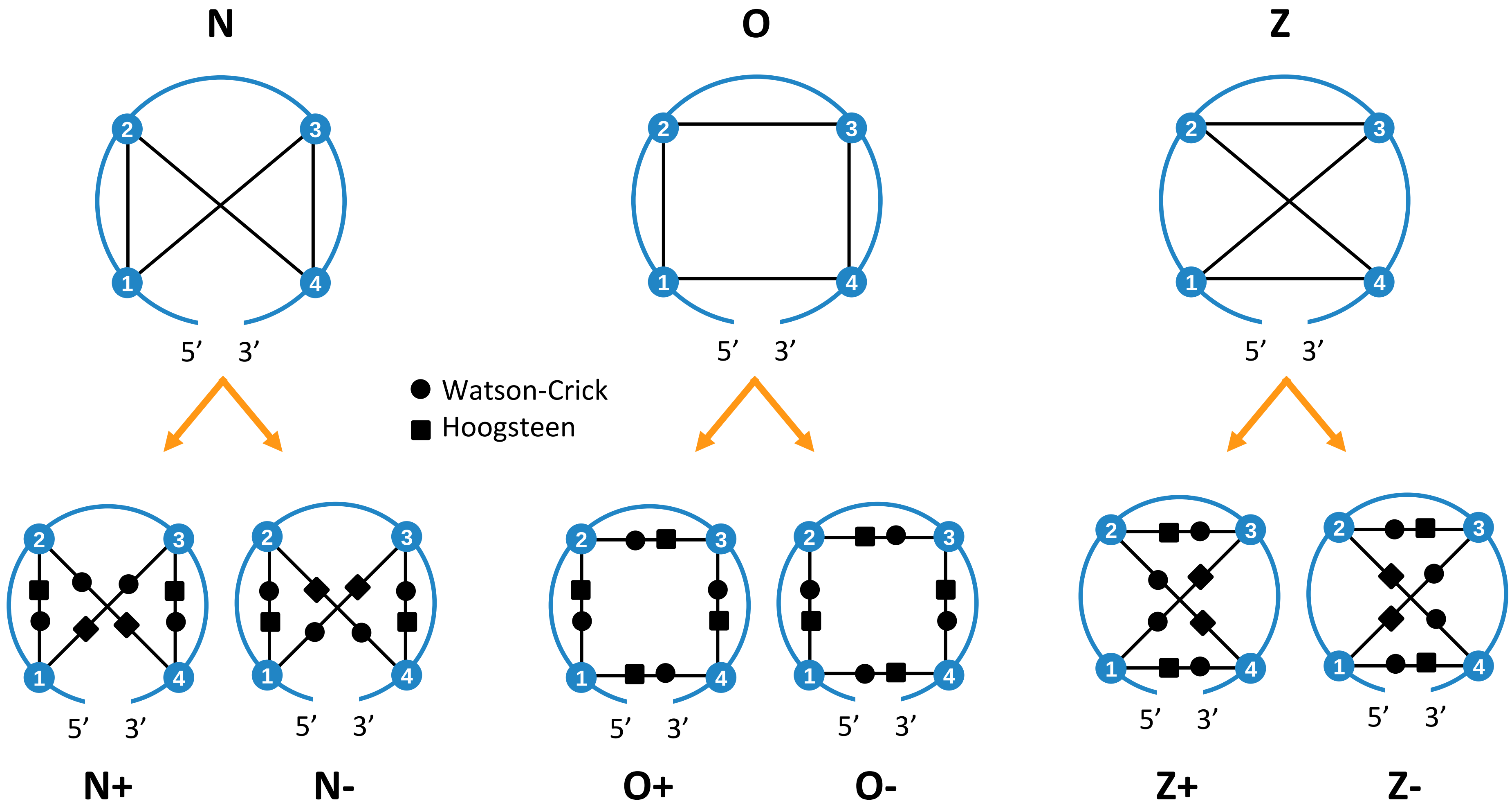
RNA Benasque, 18.08.2022

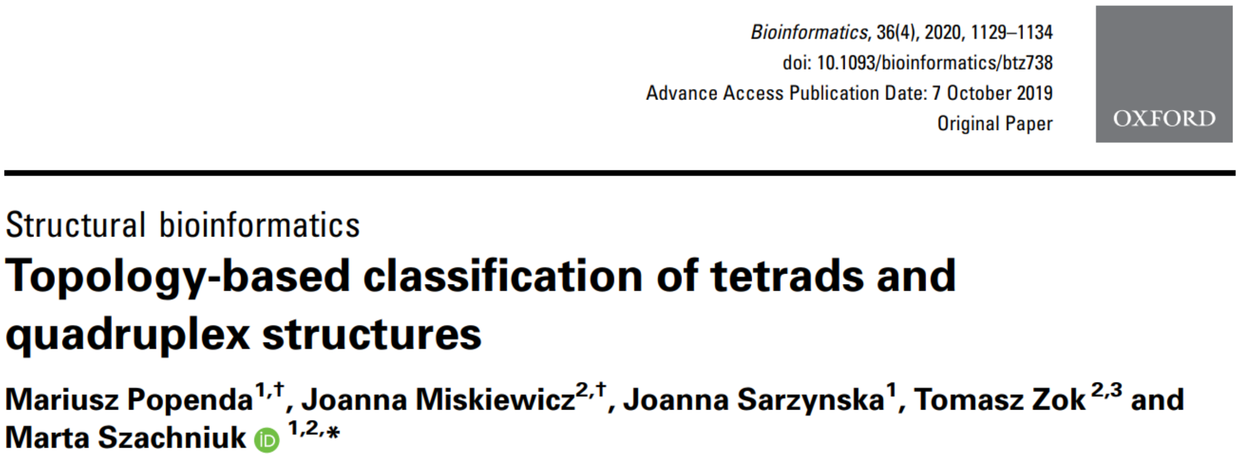
November 19, 2016


Hoogsteen - Watson-Crick
Watson-Crick - Hoogsteen

2D STRUCTURE DIAGRAMS → ONZ CLASSIFICATION

cyclic pairing schemes



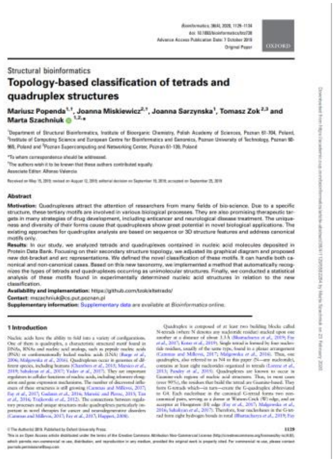


Bioinformatics, 36(4), 2020, 1129–1134
doi: 10.1093/bioinformatics/btz738
Advance Access Publication Date: 7 October 2019
Original Paper

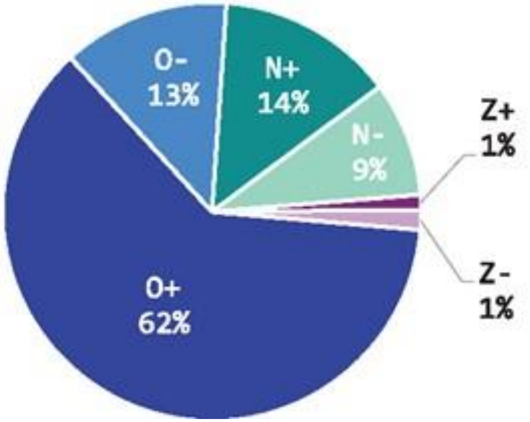
Structural bioinformatics

Topology-based classification of tetrads and quadruplex structures

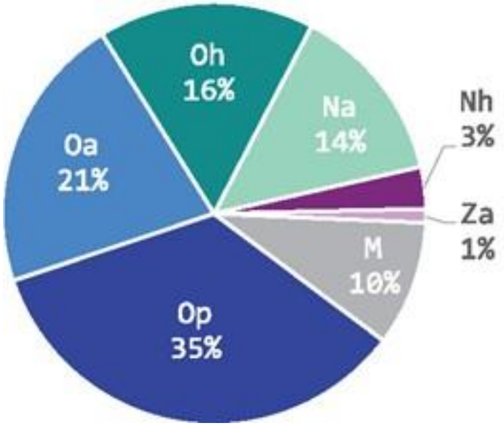
Mariusz Popenda[1,†], Joanna Miskiewicz[2,†], Joanna Sarzynska[1], Tomasz Zok[2,3] and Marta Szachniuk[1,2,*]



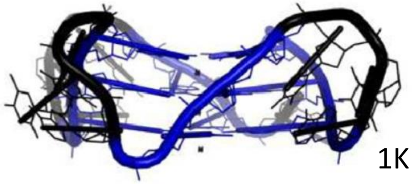
ONZ class coverage by tetrads from unimolecular quadruplexes



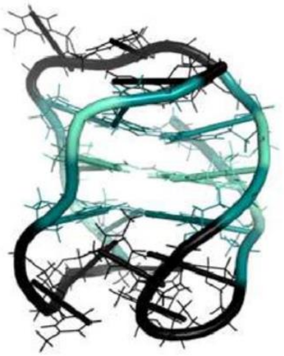
ONZM class coverage by unimolecular quadruplexes



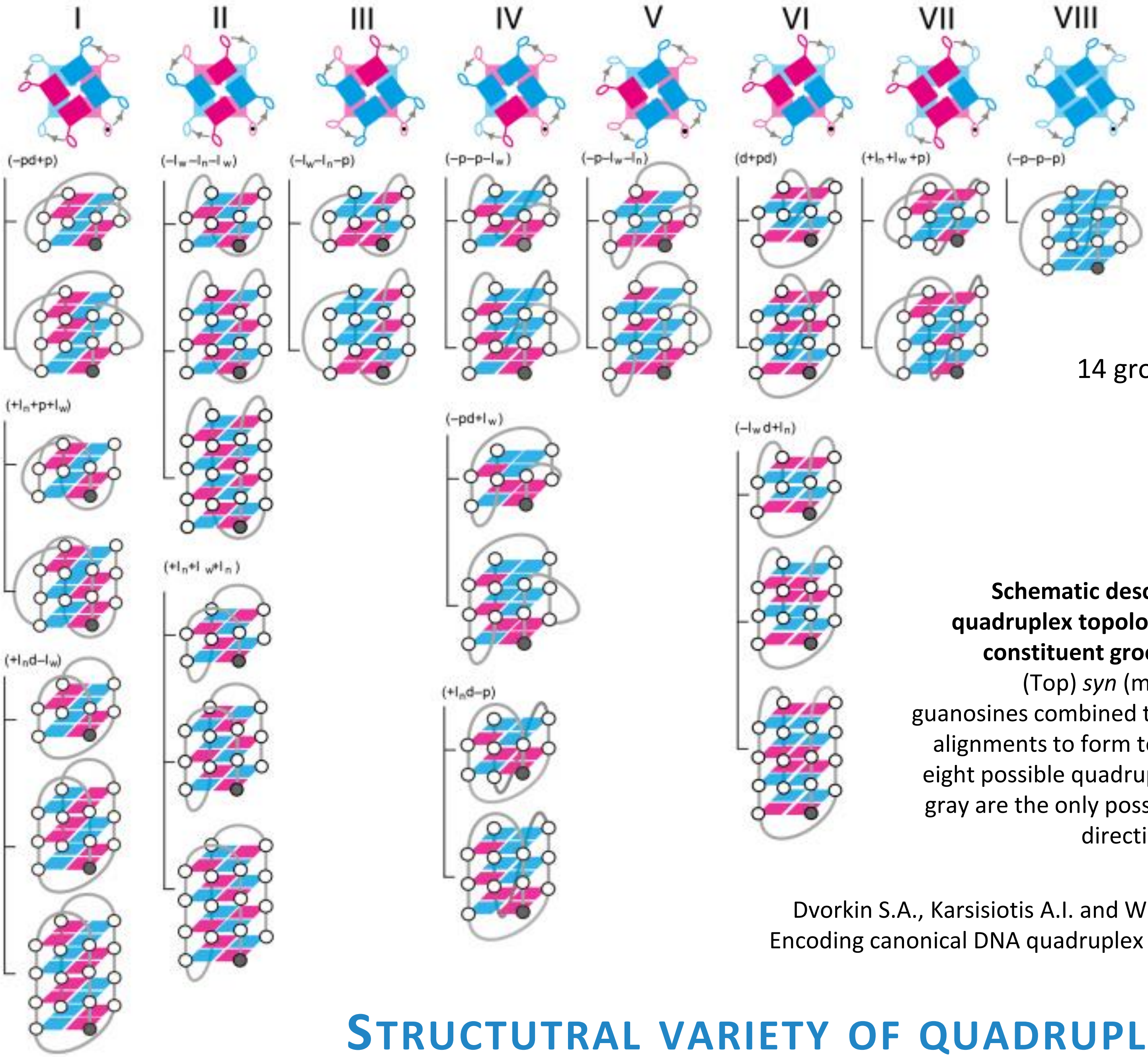
AGGGTTAGGGTTAGGGTTAGGG



1KF1



143D



8 general classes
14 groups in these classes

Schematic description of all canonical quadruplex topologies feasible with their constituent groove type combination. (Top) *syn* (magenta) and *anti* (cyan) guanosines combined through hydrogen bond alignments to form tetrads representing the eight possible quadruplex stems. Indicated in gray are the only possible strand progression directionalities for each stem.

Dvorkin S.A., Karsisiotis A.I. and Webba da Silva M. (2018) Encoding canonical DNA quadruplex structure. Sci Adv, 4(8).

STRUCTUTRAL VARIETY OF QUADRUPLEXES

DESCRIBING QUADRUPLEX STRUCTURE

- Loops (lateral, diagonal, propeller)
- Paralel/antiparallel directionality
- Right-handed/left-handed torsion
- Number of tetrads
- Sequence
- G-tracts
- Number of strands
- Rise and twist parameters
- Groove
- Planarity deviation
- Webba da Silva Classification
- ONZ classification (by our group)
- and more…

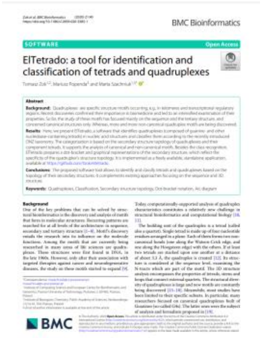

Zok *et al. BMC Bioinformatics* (2020) 21:40
https://doi.org/10.1186/s12859-020-3385-1

BMC Bioinformatics

SOFTWARE Open Access

ElTetrado: a tool for identification and classification of tetrads and quadruplexes

Tomasz Zok[1,2], Mariusz Popenda[3] and Marta Szachniuk[1,3]*

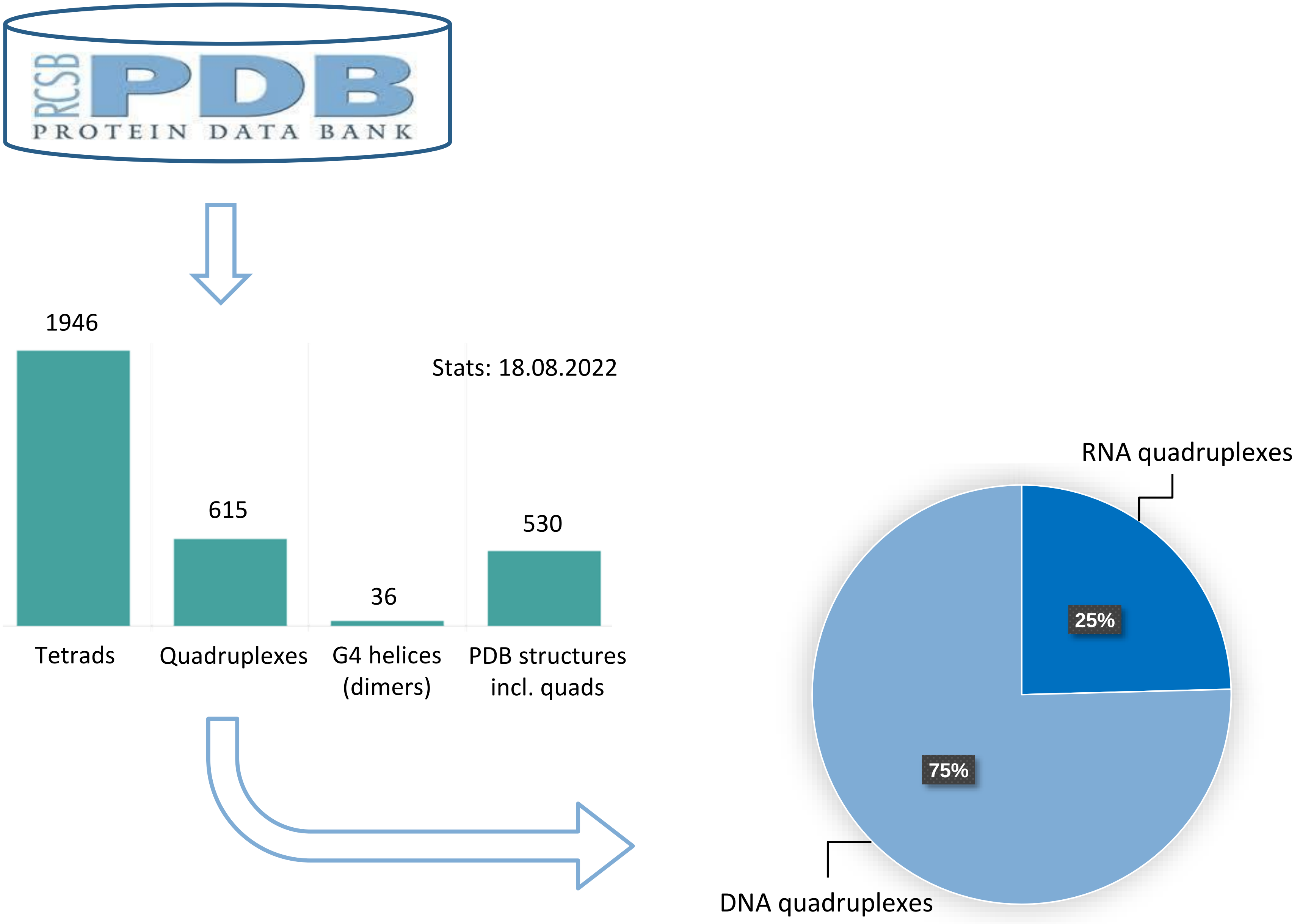


rG4 AND ITS 2D/3D STRUCTURE PROCESSING

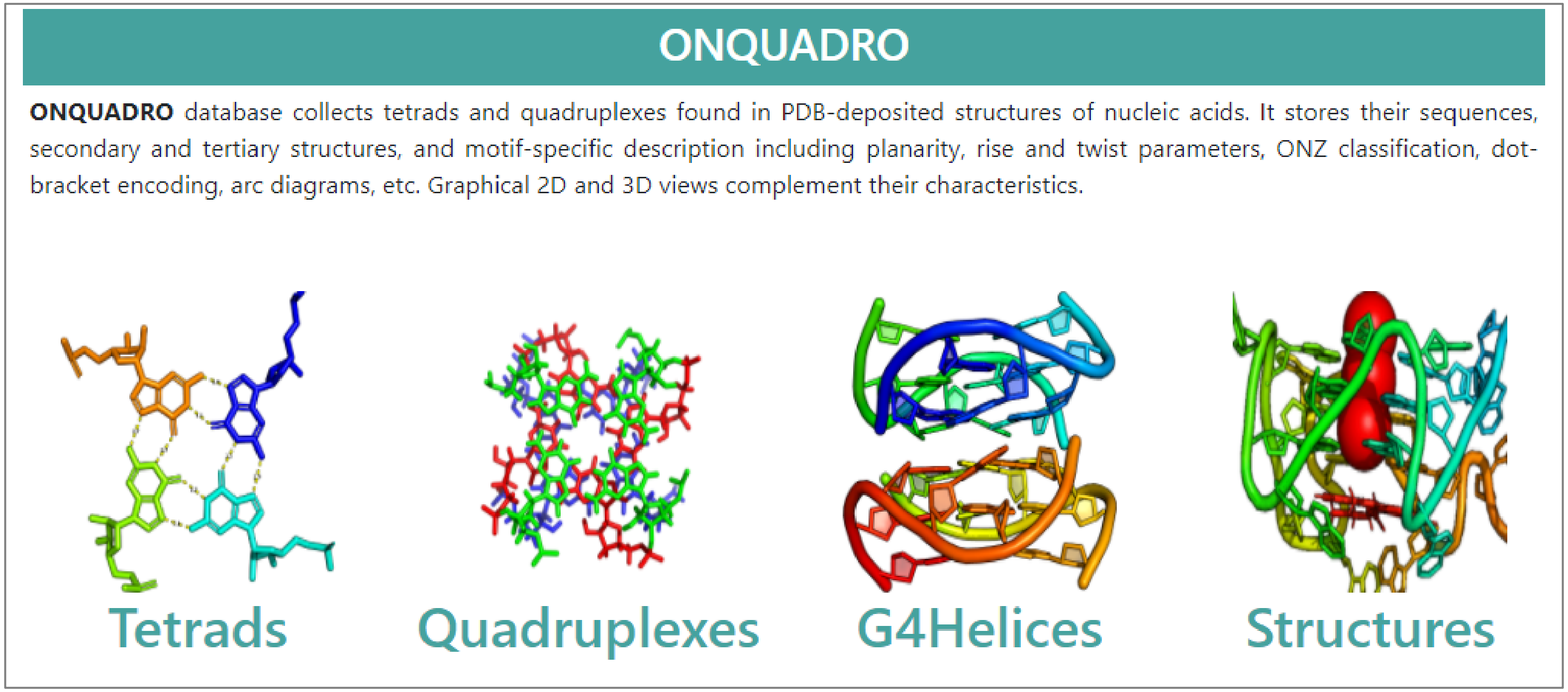
	classification		computation of structure parameters						visualization		
	G4 Classification	Base-pair Classification	Strand polarity	Area	Rise	Twist	Planarity	Torsion angles	2D view	3D view	Moving camera
DSSR	Loop-based	Saenger, LW	+	+	+	+	+	+	+	+	
ElTetrado	ONZ	LW	+		+	+	+		+		
RNApdbee		Saenger, LW							+		
3D-Nus	Q1-Q17				+	+		+		+	+

DSSR	http://skmatic.x3dna.org/
ElTetrado	https://github.com/tzok/eltetrado/
RNApdbee	http://rnapdbee.cs.put.poznan.pl/
3D-Nus	http://iith.ac.in/3dnus/

LW: Leontis-Westhof classification


RCSB PDB
PROTEIN DATA BANK
1946
615
36
530
Stats: 18.08.2022
Tetrads
Quadruplexes
G4 helices (dimers)
PDB structures incl. quads
RNA quadruplexes
25%
75%
DNA quadruplexes

https://onquadro.cs.put.poznan.pl



ONQUADRO

ONQUADRO database collects tetrads and quadruplexes found in PDB-deposited structures of nucleic acids. It stores their sequences, secondary and tertiary structures, and motif-specific description including planarity, rise and twist parameters, ONZ classification, dot-bracket encoding, arc diagrams, etc. Graphical 2D and 3D views complement their characteristics.

Tetrads

Quadruplexes

G4Helices

Structures

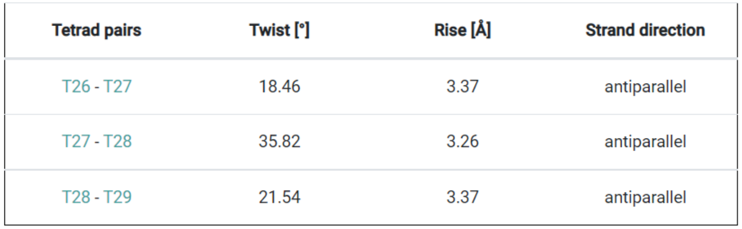
Published online 19 November 2021 *Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D253–D258***
https://doi.org/10.1093/nar/gkab1118

ONQUADRO: a database of experimentally determined quadruplex structures

Tomasz Zok [1], Natalia Kraszewska [1], Joanna Miskiewicz [1], Paulina Pielacinska [1], Michal Zurkowski [1] and Marta Szachniuk [1,2,*]

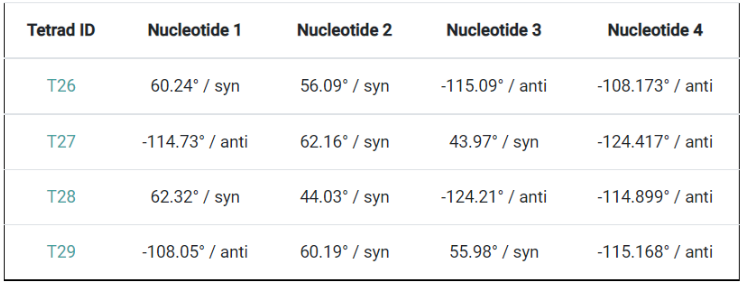
ONQUADRO

structure parameters



Tetrad pairs	Twist [°]	Rise [Å]	Strand direction
T26 - T27	18.46	3.37	antiparallel
T27 - T28	35.82	3.26	antiparallel
T28 - T29	21.54	3.37	antiparallel

Loop ID	Short sequence	Full sequence	Loop length	Loop type
L1	TTTT	A.DT5, A.DT6, A.DT7, A.DT8	4	diagonal
L2	TTTT	B.DT17, B.DT18, B.DT19, B.DT20	4	diagonal

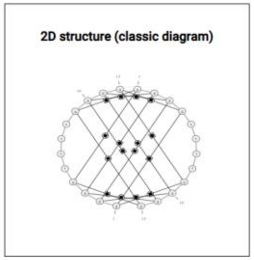


Tetrad ID	Nucleotide 1	Nucleotide 2	Nucleotide 3	Nucleotide 4
T26	60.24° / syn	56.09° / syn	-115.09° / anti	-108.173° / anti
T27	-114.73° / anti	62.16° / syn	43.97° / syn	-124.417° / anti
T28	62.32° / syn	44.03° / syn	-124.21° / anti	-114.899° / anti
T29	-108.05° / anti	60.19° / syn	55.98° / syn	-115.168° / anti

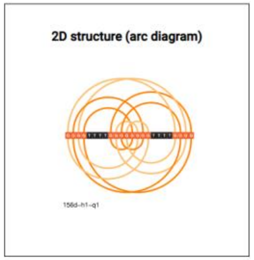
structure visualizations, dot-brackets

2D structure (dot-bracket):

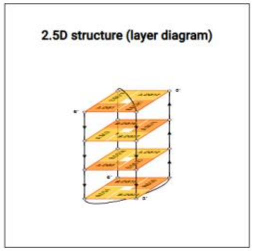
```
GGGGTTTTGGGG-GGGGTTTTGGGG
(([(....{{[[-)})]....)]]}
({([....([[{-))])....}}]]
```



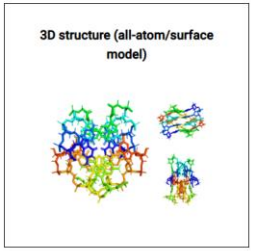
2D structure (classic diagram)



2D structure (arc diagram)

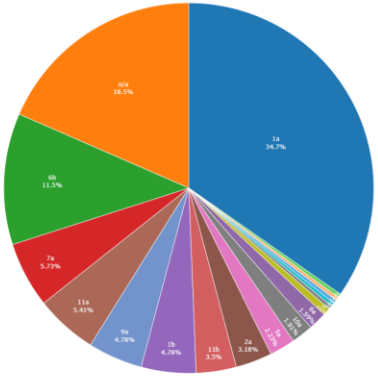


2.5D structure (layer diagram)

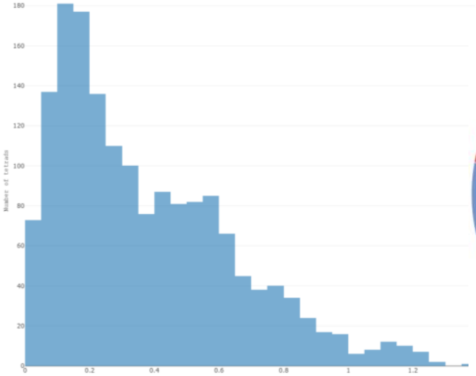


3D structure (all-atom/surface model)

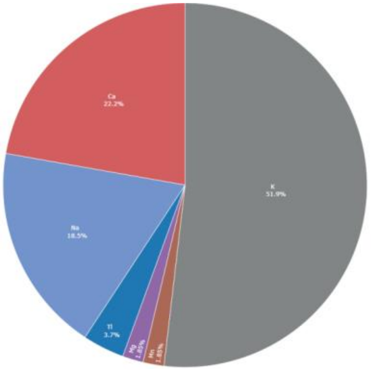
statistics



Webba da Silva classes



planarity



ions

OTHER QUADRUPLEX-RELATED DATABASES (16)

	DNA	RNA	G4 verification	G4 sequence	No. of G4s	DB records	Customized search	Web interface	Visual output
DSSR-G4DB	+	+	Experimental	+	354	354 (PDBs)	+	+	+
G4IPDB	+	+	Experimental	+	unknown	216	+	+	+
G4LDB	+	+	Experimental	+	unknown	>800	+	+	+
G4RNA	+	+	Experimental	+	321	567	+	+	
Lit392	+	+	Experimental	+	298	392			
Lit638	+	+	Experimental	+	506	638			
GSE63874	+		Experimental		716 310	32 mln			
GSE77282		+	Experimental		3 383	1,15 bln			
GSE110582	+		Experimental		1 420 841	767,39 Mb			
GSE129281		+	Experimental		329	3 505			
Greglist	+		In silico	+	Unknown	115 442	+	+	+
GRSDB2		+	In silico	+	3 255 075	29 288	+	+	+
G4-virus	+	+	In silico		47	248		+	+
Non-B DB	+		In silico	+	3 864 596	12	+	+	+
PlantGQ	+		In silico	+	626 341 645	195	+	+	+
Quadbase2	+		In silico	+	unknown	1 897	+	+	+

Briefings in Bioinformatics

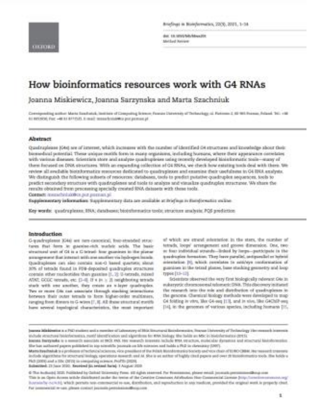
JOURNAL ARTICLE

How bioinformatics resources work with G4 RNAs

Joanna Miskiewicz, Joanna Sarzynska, Marta Szachniuk

Briefings in Bioinformatics, Volume 22, Issue 3, May 2021, bbaa201, https://doi.org/10.1093/bib/bbaa201

Published: 08 September 2020 **Article history** ▾



How bioinformatics resources work with G4 RNAs

Joanna Miskiewicz, Joanna Sarzynska and Marta Szachniuk

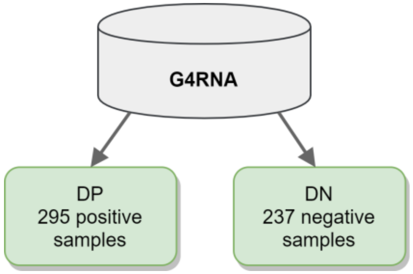
Sequence analysis and prediction of sites where quadruplexes can form (14)	G4Hunter G4Catchall G4PromFinder G4-iM Grinder G4-Predictor ImGQfinder G4Predict Pqsfinder Quadron G4P Calculator G4RNA Screener QGRS Mapper QPARSE TetraplexFinder
Secondary structure prediction (1)	RNAfold
Analysis of 2D & 3D structure (4)	DSSR RNApdbee ElTetrado 3D-Nus
Databases (16)	G4LDB G4RNA G4IPDB G4-virus GRSDB2 Greglist DSSR-G4DB Lit392 Lit638 Quadbase2 Plant-GG Non-B DB GSE63874 GSE77282 GSE110582 GSE129281

G4 PATTERNS IN PQS PREDICTION ALGORITHMS

Tool	Pattern	Pattern description	Exemplary motif	
G4Catchall	G_n	Perfect G-tract	GGG	n = 3
	$G_xN_yG_{(n-x)}$	Bulged G-tract	GGAG	n = 3; x = 2; y = 1
	$G_xN_yG_{(n-x-1)}$	Mismatched G-tract	GGAG	n = 4; x = 2; y = 1
G4Hunter	n/a	n/a	n/a	
G4-iM Grinder	n/a	n/a	n/a	
G4P Calculator	G_n	Perfect G-tract	GGG	n = 3
G4Predict (Intra)	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for intramolecular PQS	GGAAGGAAAGGAAGG	x1= x2= x3= x4= 2; y1= y3= 2; y2= 3
G4-Predictor V.2	$G_{2\text{-}6}N_{0\text{-}36}G_{2\text{-}6}N_{0\text{-}36}G_{2\text{-}6}N_{0\text{-}36}G_{2\text{-}6}$	Patterns depend on users' settings	GGAGGGAAGGAAGGG	
G4PromFinder	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for PQS	GGAGGGAAGGAAGGG	x1= x3= 2; x2= x4= 3; y1= 1; y2= y3= 2
G4RNA screener	n/a (ML algorithm)	n/a	n/a	
ImGQfinder	G_n	Perfect G-tract	GGG	n = 3
	$G_{i-1}NG_{n-i+1}$	Bulged G-tract	GGAGG	n = 4; i = 3
	$G_{i-1}NG_{n-i}$	Mismatched G-tract	GGAG	n = 4; i = 3
pqsfinder	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for PQS	GGAGGAAGGAAGGG	x1= x2= x3= 2; x4= 3; y1= 1; y2= y3= 2
QGRS Mapper	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for PQS	GGAGGAAGGAAGG	x1= x2= x3= x4= 2; y1= 1; y2= y3= 2
QPARSE	n/a	Patterns depend on users' settings	n/a	
Quadron	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for PQS	GGAGGAAGGAAGG	x1= x2= x3= x4= 2; y1= 1; y2= y3= 2
TetraplexFinder	$G_{x1}N_{y1}G_{x2}N_{y2}G_{x3}N_{y3}G_{x4}$	General pattern for PQS	GGAAAGGAAGGAAGG	x1= x2= x3= x4= 2; y1= 3; y2= y3= 2

BENCHMARKING ALGORITHMS FOR G4 LOCATION PREDICTION

tested for RNA quadruplexes



G4RNA

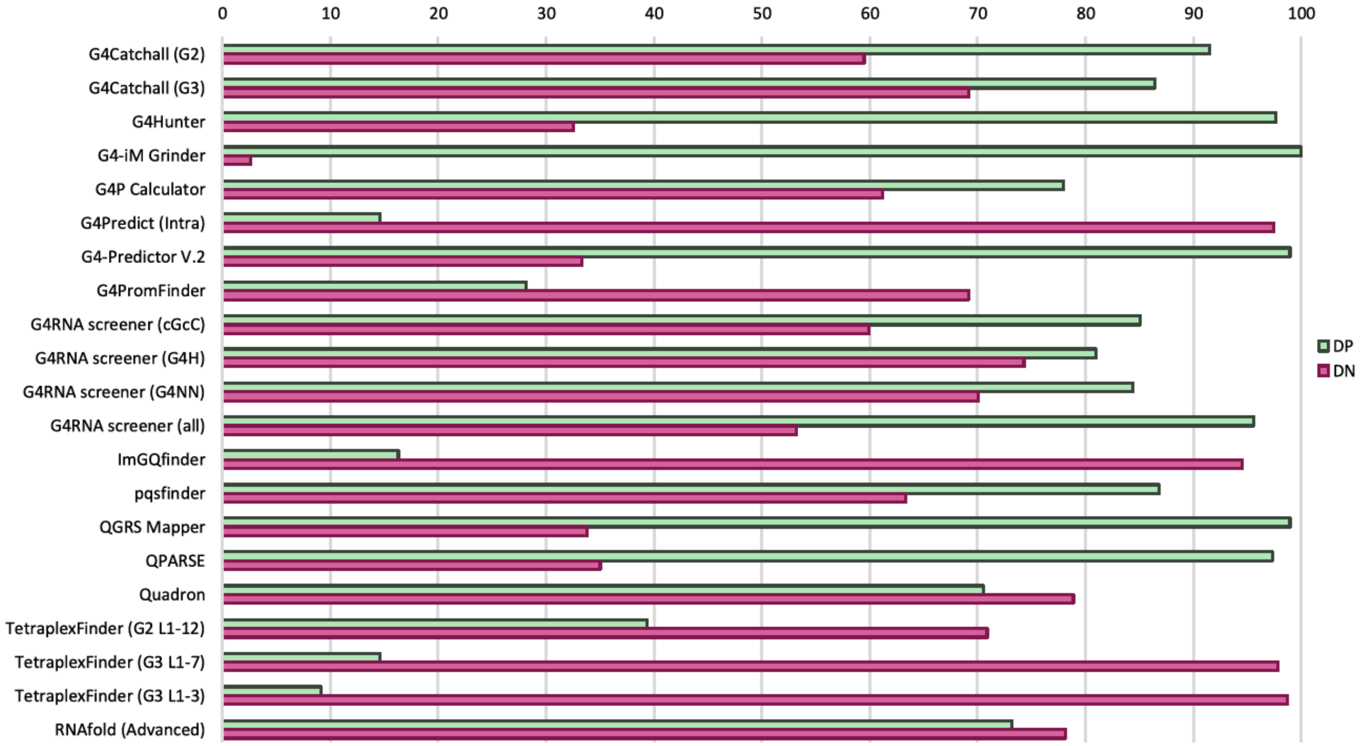
DP
295 positive samples

DN
237 negative samples

Positive (DP) and negative (DN) datasets

Best results:
G4RNA screener
G4Catchall
RNAfold

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



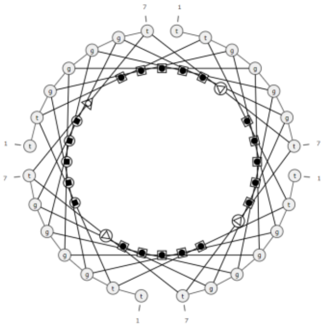
BIG CHALLENGES - PREDICTION

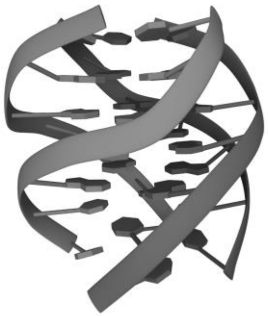
TTGGGGTAACACCTTG
GGGTGCCTTGGGGTAT
ATACCCCCCTTGGGG

TTGGGGTAACACC**TTG**
GGGTGCC**TTGGGGT**AT
ATACCCCCC**TTGGGG**

- Predict the location of non-G-quads
- Predict the secondary structure

  (non-canonical bps!)
- Predict the 3D structure of quadruplexes





Marta Szachniuk

Tomasz Żok

Joanna Miśkiewicz

Michał Żurkowski

Natalia Kraszewska

Mariusz Popenda

Joanna Sarzyńska

Take home message ☺