



In silico exploration of quadruplex structures

Joanna Miśkiewicz (1), Mariusz Popenda (2), Joanna Sarzyńska (2), Tomasz Żok (1,3), Paulina Pielacińska (1), Natalia Kraszewska (1), Marta Szachniuk (1,2)

Corresponding authors: jmiskiewicz@cs.put.poznan.pl, mszachniuk@cs.put.poznan.pl

(1) Institute of Computing Science, Poznan University of Technology, Poland
 (2) Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poland
 (3) Poznan Supercomputing and Networking Center, Poland

Introduction

G-quadruplexes (G4s) are structural motifs that appear in the DNA and RNA of many organisms. They are built of tetrads stacked upon one another. Four guanines in a pseudo-planar arrangement, connected by hydrogen bonds, form a tetrad. G4s are involved in many biological processes, for example, transcription regulation and genome stabilization. Thus, they constitute an interesting target of novel therapeutic designs.

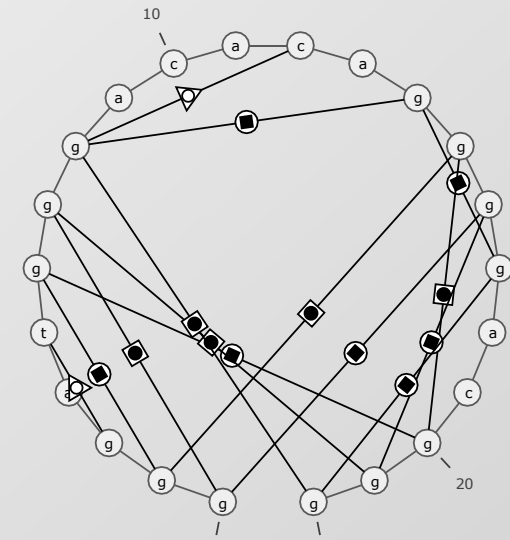
Quadruplex structures

I GGGATGGGACACAGXGGACGGG

Motif: $G_{x_1}N_{y_1}G_{x_2}N_{y_2}G_{x_3}N_{y_3}G_{x_4}$
 $x \geq 2, y \geq 0$

Quadruplex structures

II
 ((...)[(.....)](.....))
 (.....(.....))



n4-helix with 3 tetrads
 Mh* quadruplex with 3 tetrads
 A.DG1 A.DG16 A.DG21 A.DG7 cWH-cWH-cWH-cWH N-
 A.DG2 A.DG6 A.DG20 A.GF2/15 cWH-cWH-cWH-cWH N+
 A.DG8 A.GFL14 A.DG17 A.DG22 cWH-cWH-cWH-cWH O+

This work was supported by the Polish National Science Centre, grant number 2019/35/B/ST6/03074.

ONZ classification

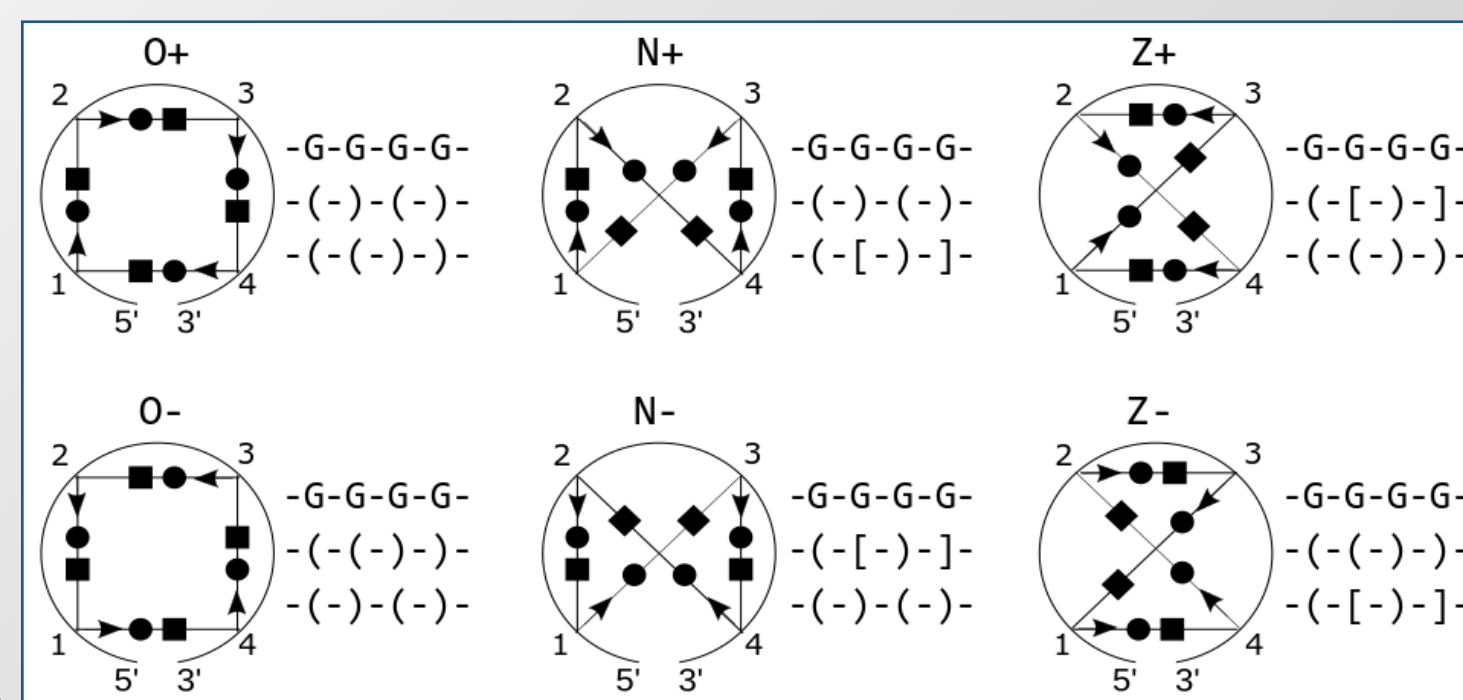
The secondary structure of tetrad T can be represented as cyclic graph $G = (V, E)$, where $|V| = |E| = 4$, each $v \in V$ represents one nucleotide from the tetrad, every $e \in E$ corresponds to a hydrogen-bonding interaction between respective nucleotides. If we placed the Vertices of G at equal distances on a circle clockwise, in the order imposed by the sequence, we'd see that graph takes the shape of a square (O-shaped), a bow tie (N-shaped), or an hourglass (Z-shaped). This observation made us distinguish 3 groups of tetrads and define their ONZ taxonomy:

Let T denote a tetrad build of N_1, N_2, N_3, N_4 nucleotides. We define ONZ classes:

- Class O if $T = \{(N_1, N_2), (N_2, N_3), (N_3, N_4), (N_4, N_1)\}$,
- Class N if $T = \{(N_1, N_2), (N_2, N_4), (N_4, N_3), (N_3, N_1)\}$,
- Class Z if $T = \{(N_1, N_3), (N_3, N_2), (N_2, N_4), (N_4, N_1)\}$.

ONZ classification

We can annotate the tetrad according to the interaction arrangement. If the first nucleobase binds with the next one along the Watson-Crick edge, the tetrad is tagged positive (+), otherwise, it is tagged negative (-).

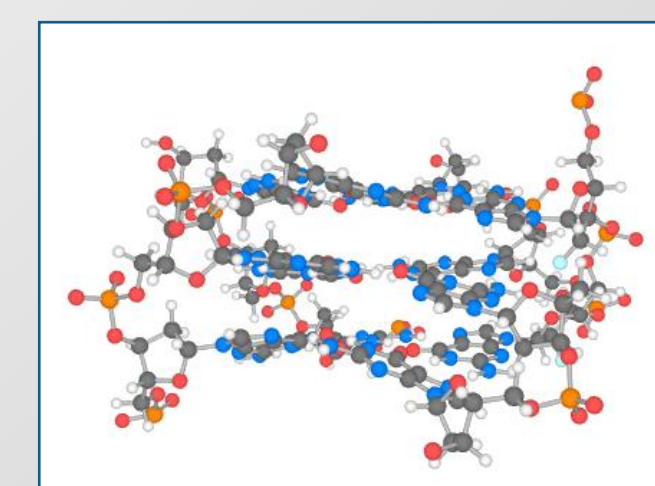


Thus, every class in ONZ is divided into two subcategories: O+, O-, N+, N-, Z+, Z-.

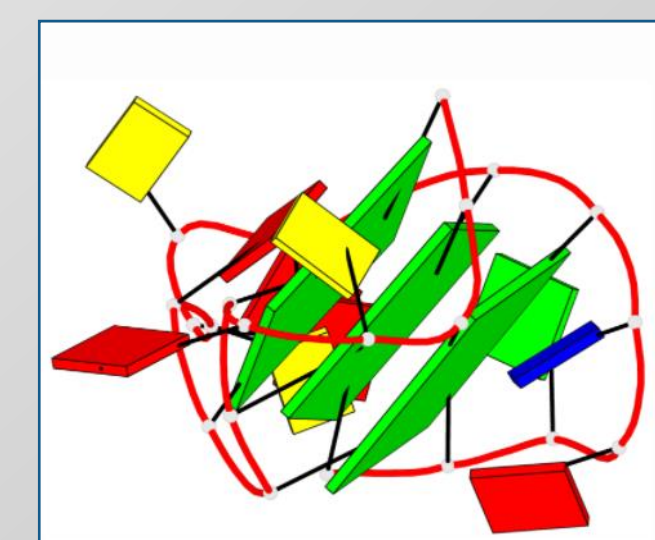
Quadruplex structures

III

Tetrad	ONZ class	Planarity
1	N-	0.17
2	N+	0.34
3	O+	0.10



Base pair	Twist	Rise
1 - 2	23.43	3.45
2 - 3	56.18	6.79



Analysed structure: 6TC8

ElTetrado

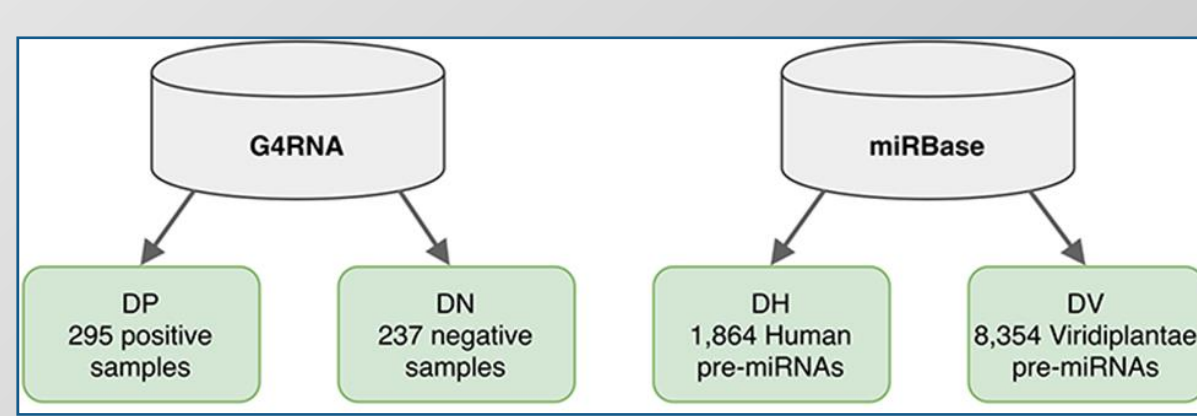
ElTetrado identifies and describes tetrads and quadruplexes in the 3D structures of nucleic acids, by searching for G-based and non-G-based motifs. It allocates tetrads and quadruplexes to ONZ classes according to their 2D structure topology, calculates strand direction, planarity deviation, rise and twist parameters. The program also outputs the graphical representation of the 2D structure (top-down arc diagram) and its dot-bracket encoding in a two-line format—both designed specially to handle tetrads and quadruplexes.

	Op	Oa	Oh	Na	Nh	Mp	Mh	Total
+	105	-	12	11	3	8	8	147
-	2	-	-	-	-	-	-	2
*	13	40	23	27	4	5	6	118
Total	120	40	35	38	7	13	14	267

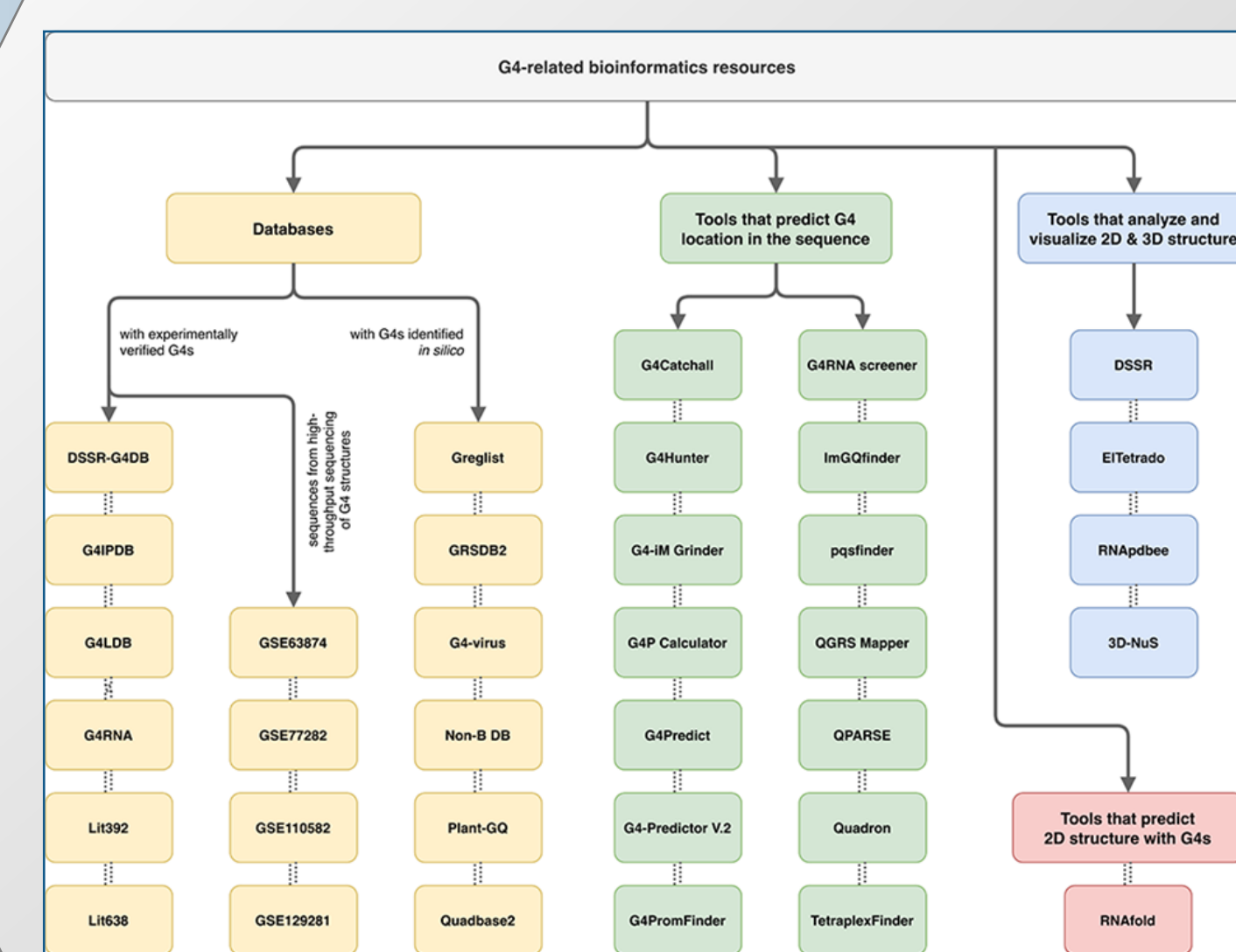
ONZM class coverage by unimolecular quadruplexes

Quadruplex resources

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 creation of G4-related databases that primarily collect information about sequences with the ability to form quadruplexes. We distinguished the following subsets of resources: databases, tools to predict putative quadruplex sequences, tools to predict secondary structure with quadruplex motifs, and tools to analyze and visualize quadruplex structures.

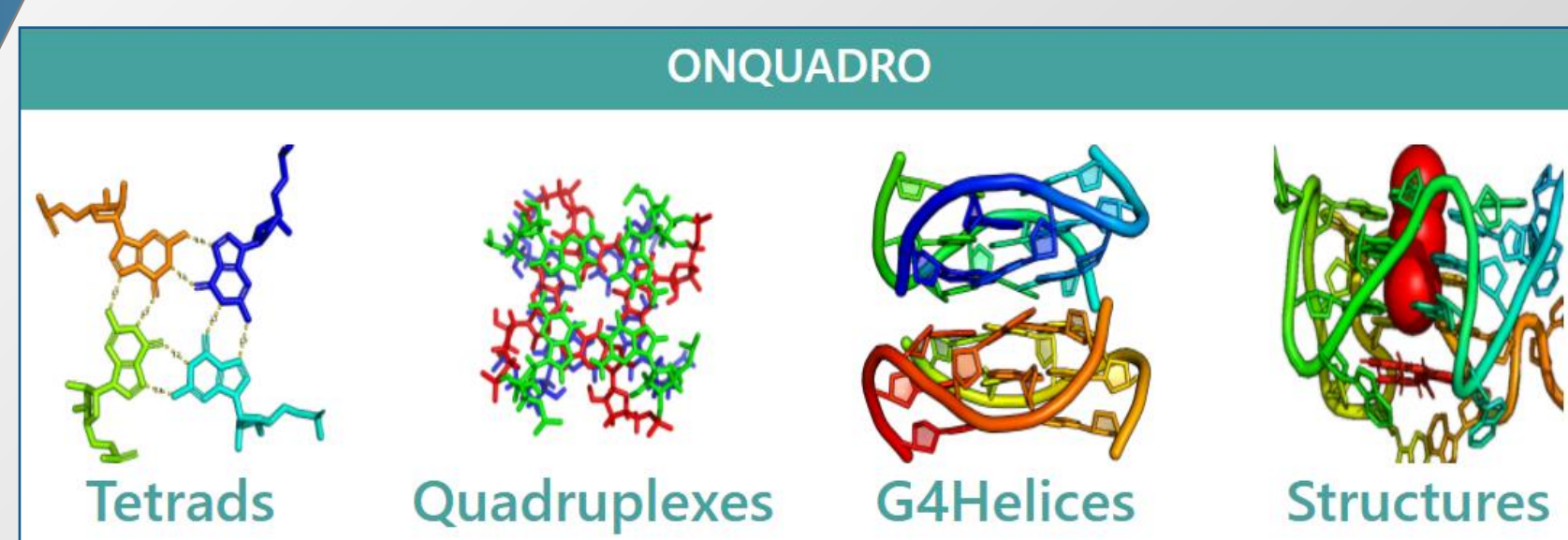


Quadruplex resources



ONQUADRO

ONQUADRO database collects tetrads and quadruplexes found in PDB-deposited structures of nucleic acids.



It stores their sequences, 2D and 3D structures, and motif-specific description including planarity, rise and twist parameters, ONZ classification, dot-bracket encoding, arc diagrams, graphical views of 2D and 3D structure, etc.

Quadruplex resources

Analysis of G4-dedicated programs



Coverage of DP and DN datasets with correct predictions: positive in DP and negative in DN [%]. The best results: G4RNA screener, G4Catchall, RNAfold.

References

- [ONZ] Popenda et al. Topology-based classification of tetrads and quadruplex structures. *Bioinformatics* 2020, 36:1129–34.
 [ElTetrado] Żok et al. ElTetrado: a tool for identification and classification of tetrads and quadruplexes. *BMC Bioinformatics* 2020, 21:40.
 [Quadruplex resources] Miśkiewicz et al. How bioinformatics resources work with G4 RNAs. *Briefings in Bioinformatics* 2020, ahead of print.
 [ONQUADRO] <http://onquadro.cs.put.poznan.pl/>

RNA junctions from a 3D structure perspective

J. Wiedemann*¹, M. Antczak*^{1,2}, J. Kaczor¹, M. Milostan^{1,3}, T. Zok¹, M. Szachniuk^{1,2}

1. Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland
2. Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
3. Poznan Supercomputing and Networking Center, Jana Pawla II 10, 61-131 Poznan, Poland



Introduction

Computational methods for the 3D structure prediction allow for prototyping the shape of RNA, yet some of its fragments require more attention and manual or semi-automatic adjustment. Among them are multibranch loops (n-way junctions) - hard to predict structural motifs that significantly impact the structure of the whole molecule.

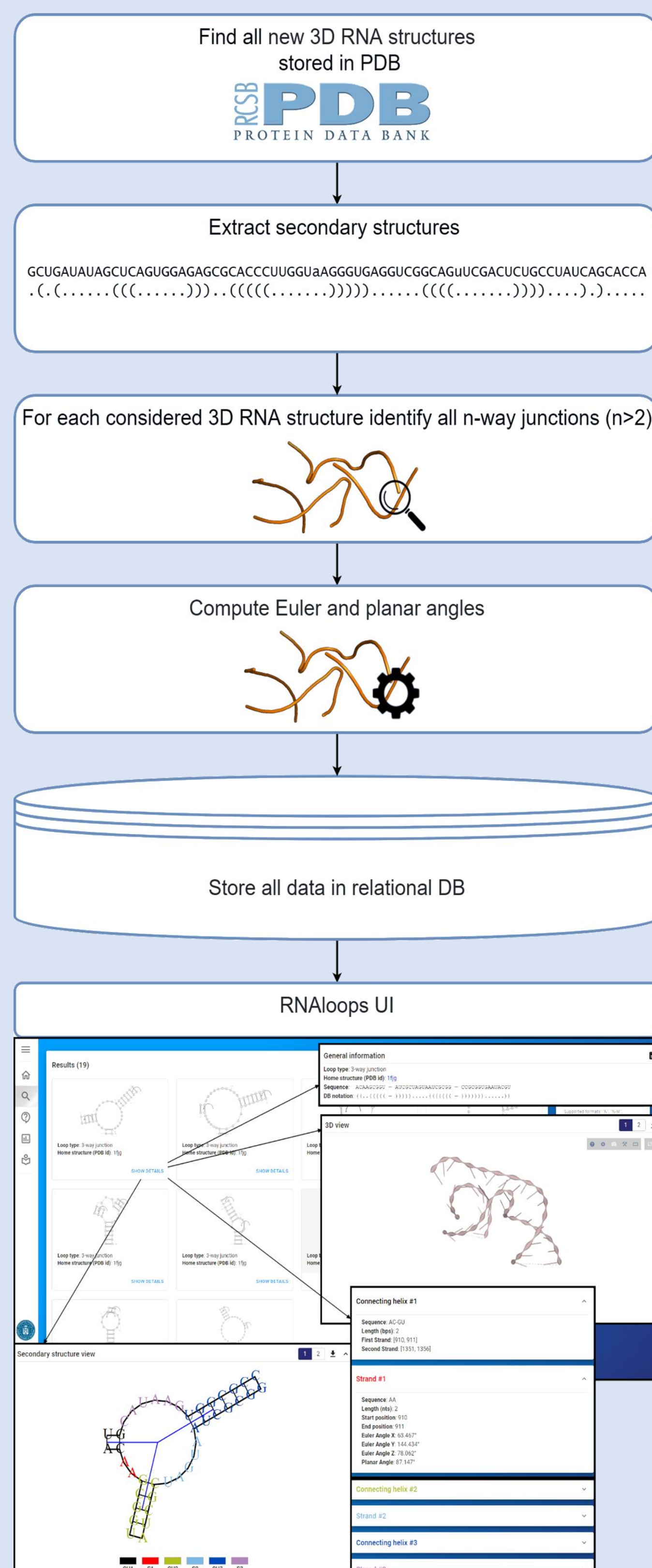
In this work, we created the RNAloops database that collects structural data of RNA n-way junctions. The novelty in our tool is the loop description that contains, i.a., a set of angles (Euler and planar) to determine spatial relationship between outgoing helices.

Data analysis performed with the RNAloops contents showed that every eight RNA contain n-way junctions; in some structures we found even >100 of them per molecule. We believe that data collected in RNAloops can be used to improve the accuracy of *in silico* modeling of RNA 3D structures.

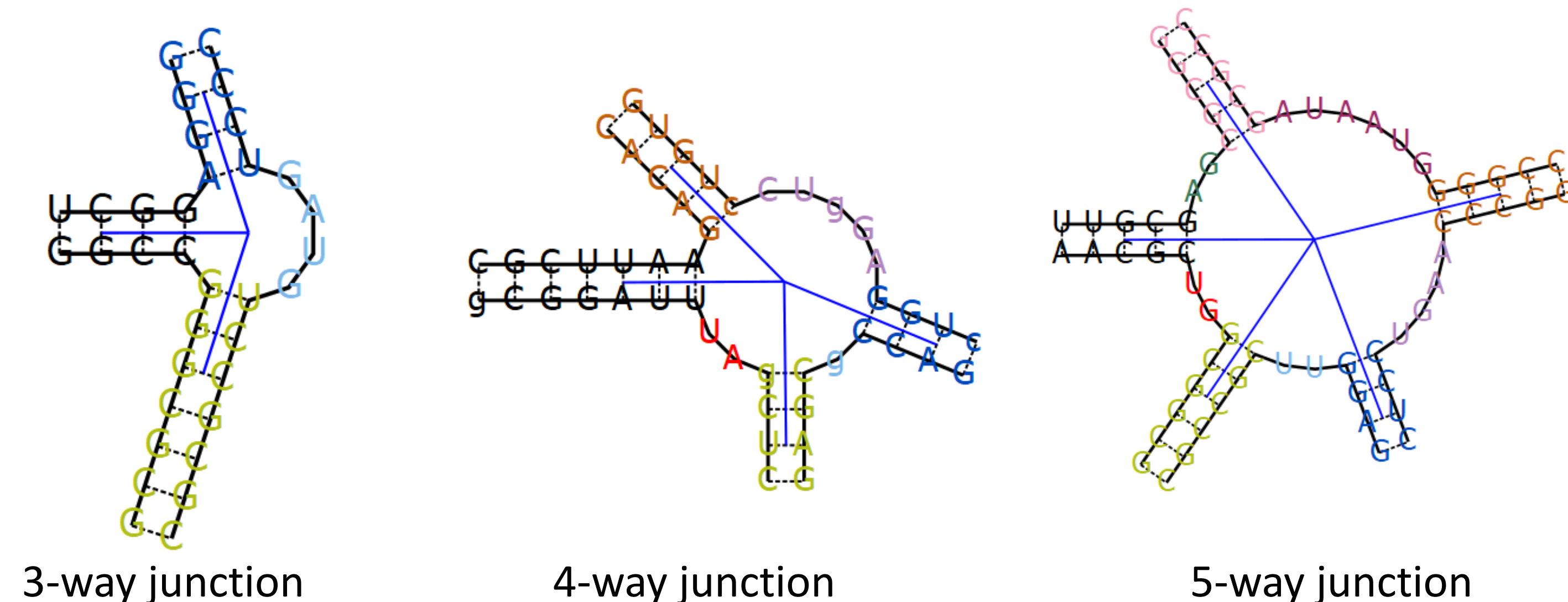
About RNAloops

RNAloops stores the information about n-way junctions found in experimentally determined RNA 3D structures deposited in the Protein Data Bank. The stored data include sequence, secondary structure, tertiary structure, one planar, and three Euler angles that describe the relationship between stems coming out of the loop.

The database is automatically updated every Sunday. Newly deposited RNA 3D structures are automatically downloaded from the RCSB PDB repository. For every downloaded RNA, its secondary structure is extracted using RNApdbee algorithm and encoded in extended dot-bracket notation. All n-way junctions are identified in each RNA, based on its secondary structure processing. Planar angle and Euler angles are computed for every pair of outgoing helices and attached to the description of the multiloop.



Example n-way junctions in RNAloops

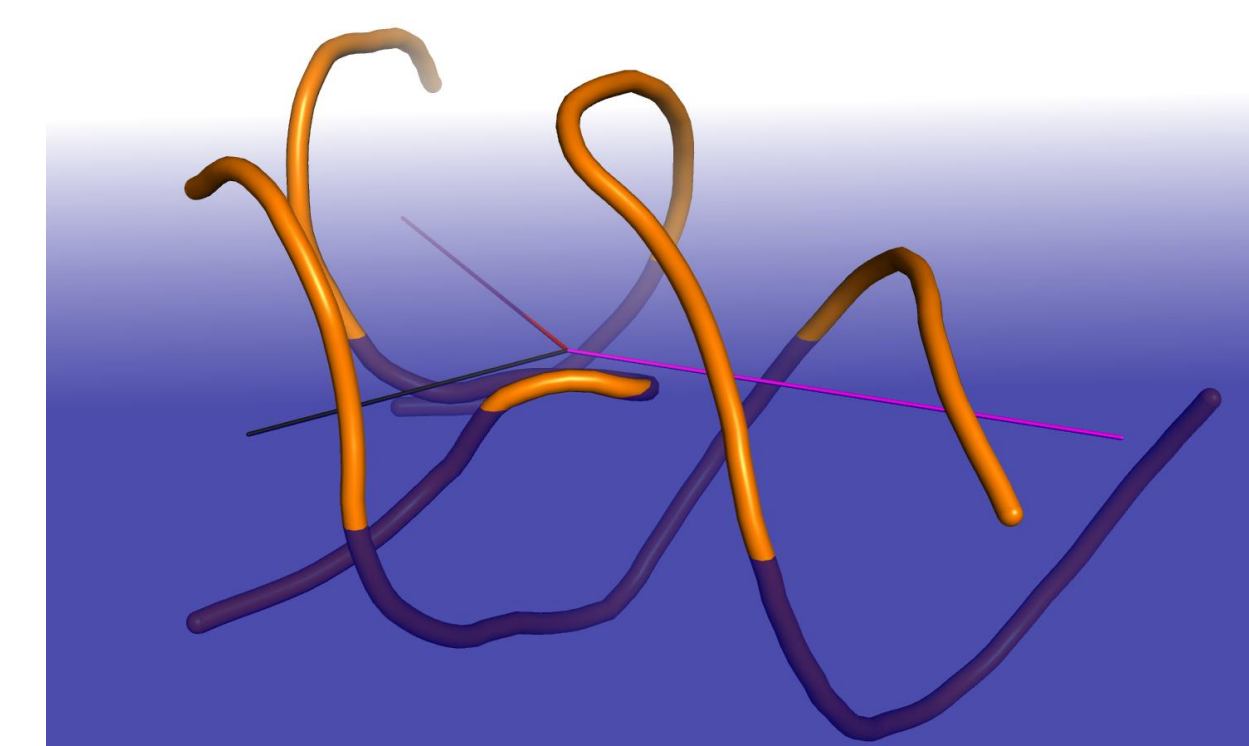


Angular representation

- **Euler angles** (X, Y, Z) – three angles describing the rotation around axes in 3D that are required to align two neighbouring connecting helices.

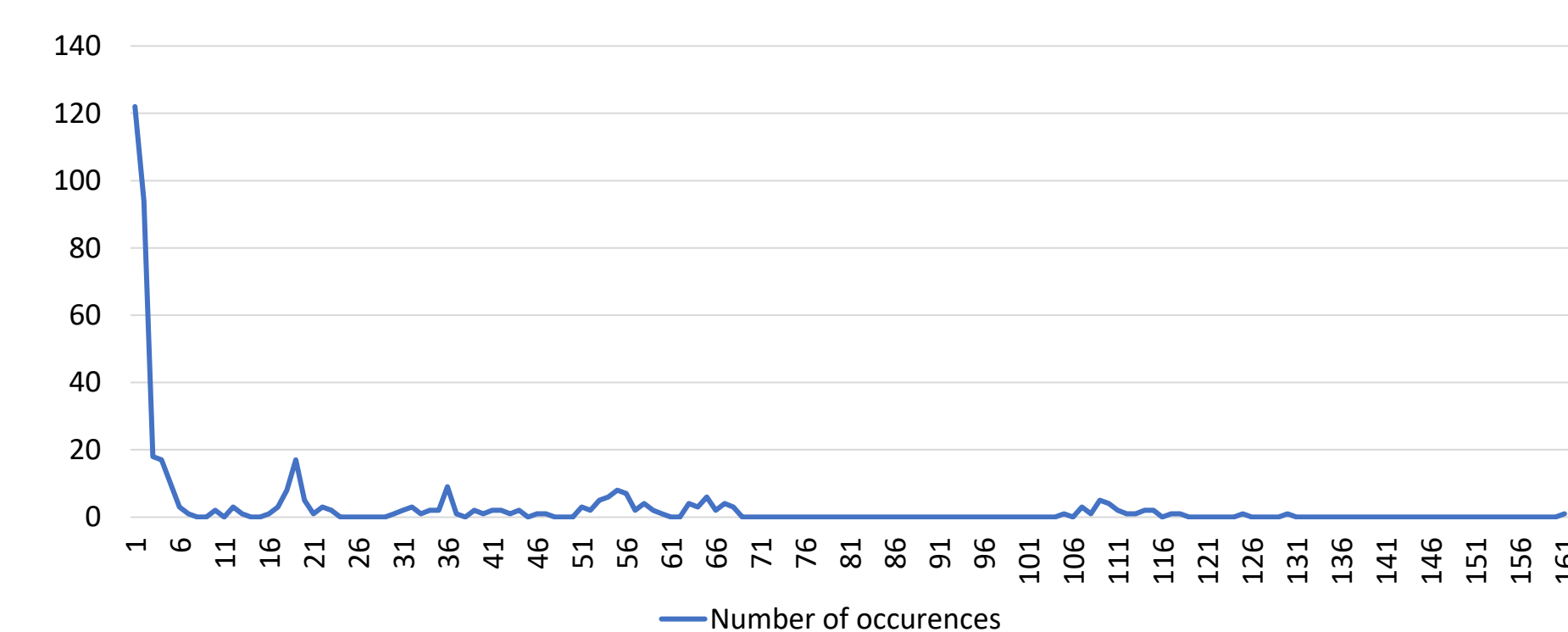
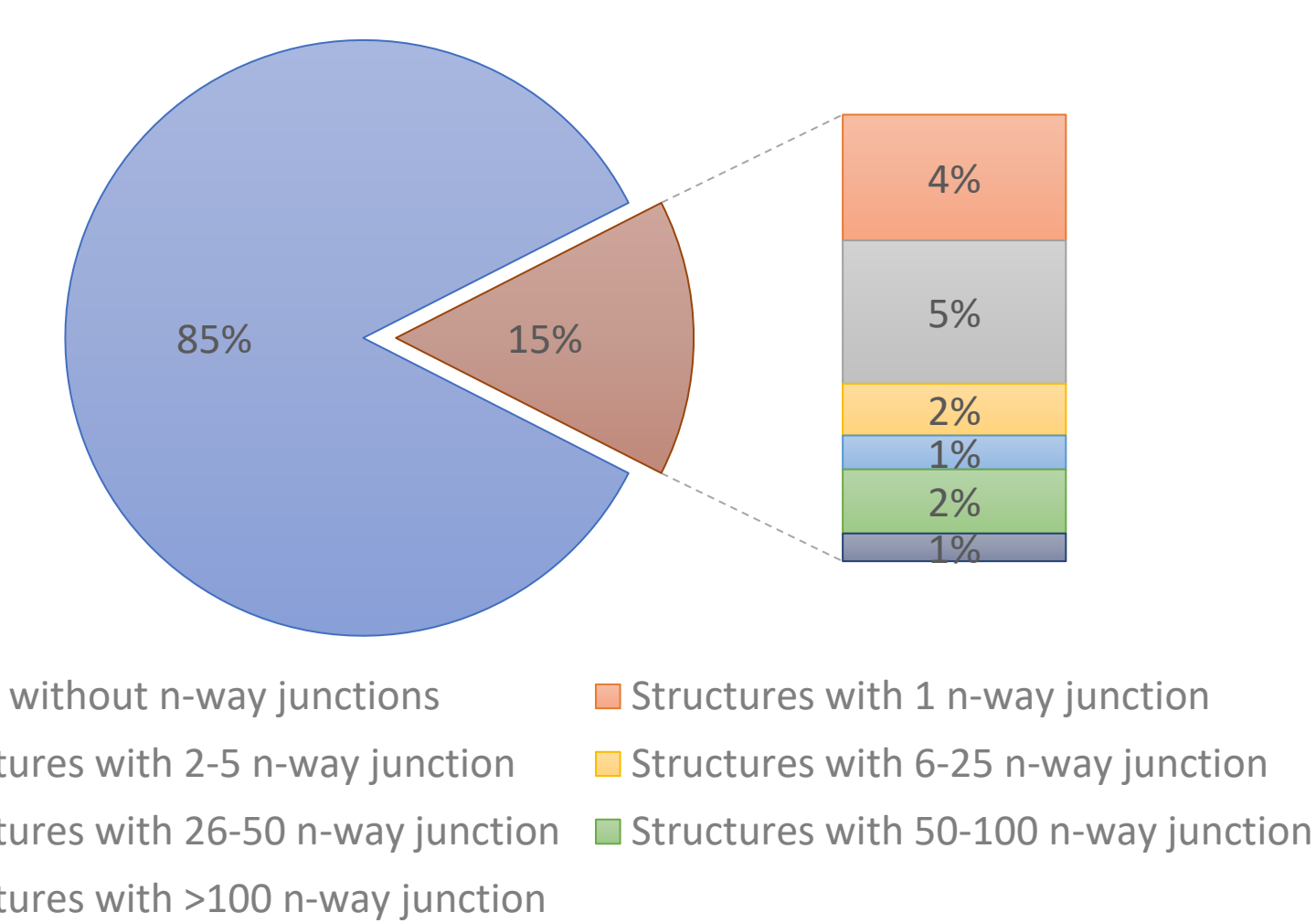


- **Planar angle** – a single angle describing angle between two neighbouring connecting helices. Figure below visualizes example of the planar angle.

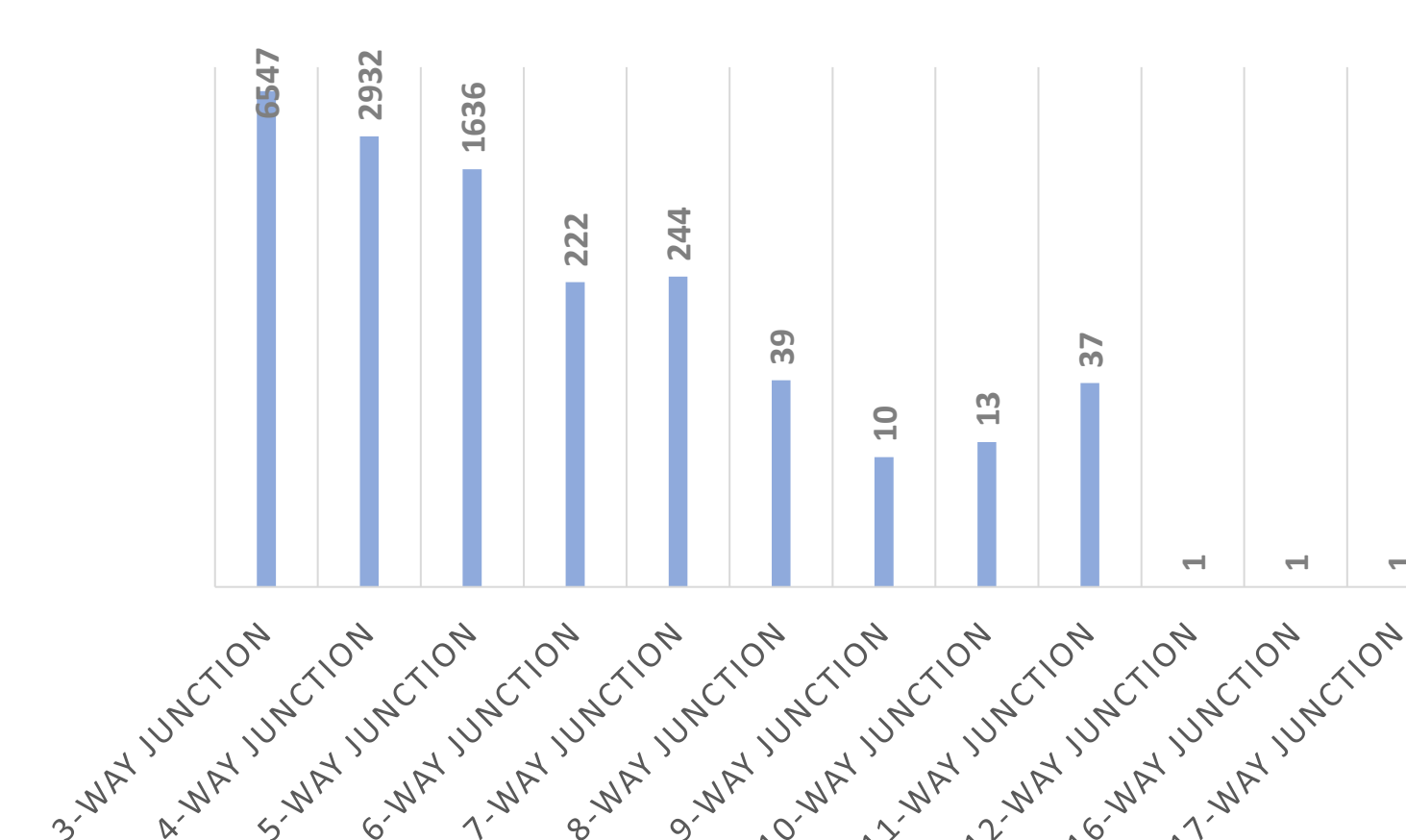


Results

The RNAloops database stores information of 11 984 junctions from 510 RNA structures, as of 3-Nov-2020. It is the result of processing 2 418 RNA structures from the RCSB PDB repository.



We identified junctions with 3-17 outgoing helices. Out of these 11 984 junctions, over 50% were 3-way junctions (6 547), about 30% were 4-way junctions, 13% - 5-way junctions.



The influence of structure size on similarity metrics values.

Piotr Kłosowicz², Tomasz Żok¹

¹ Poznan University of Technology, Poland

² Adam Mickiewicz University Poznan, Poland

Introduction

A comparison of tertiary structure between several molecules is very important in order to understand function of specific RNA molecules and interactions between them. With it we can find motifs, that are crucial to identify and recognize to the role of newly-found molecules or extend our knowledge about known RNA chains, which are still not fully studied. But in order to make a comparison, we need a measure, that is reliable and can be used to a wide variety of molecules, with as low number of limitations as possible.

Our inspiration and objectives

In order to overcome this problem, a group of scientists from University of North Carolina took a closer look at root-mean-square deviation (RMSD). Their work was described in article “On the significance of an RNA tertiary structure prediction”^[1], written in 2010. They have discovered, that RMSD is highly dependant on the size of the molecule and in order to make it more reliable, they proposed usage of prediction significance (P-value). This value allows to evaluate, if the given prediction is better than one expected by chance for molecule of that size.

But there are more measures, that can be used to compare tertiary structures than just RMSD. In our research we are concentrating on mean of the circular quantities (MCQ), which is an approach used in the MCQ4Structures tool. It is described in “MCQ4Structures to compute similarity of molecule structures”^[2] article, written in 2014 by the group of scientists from Poznan University of Technology. We have decided to take a closer look on this measurement in order to make it size-independant.

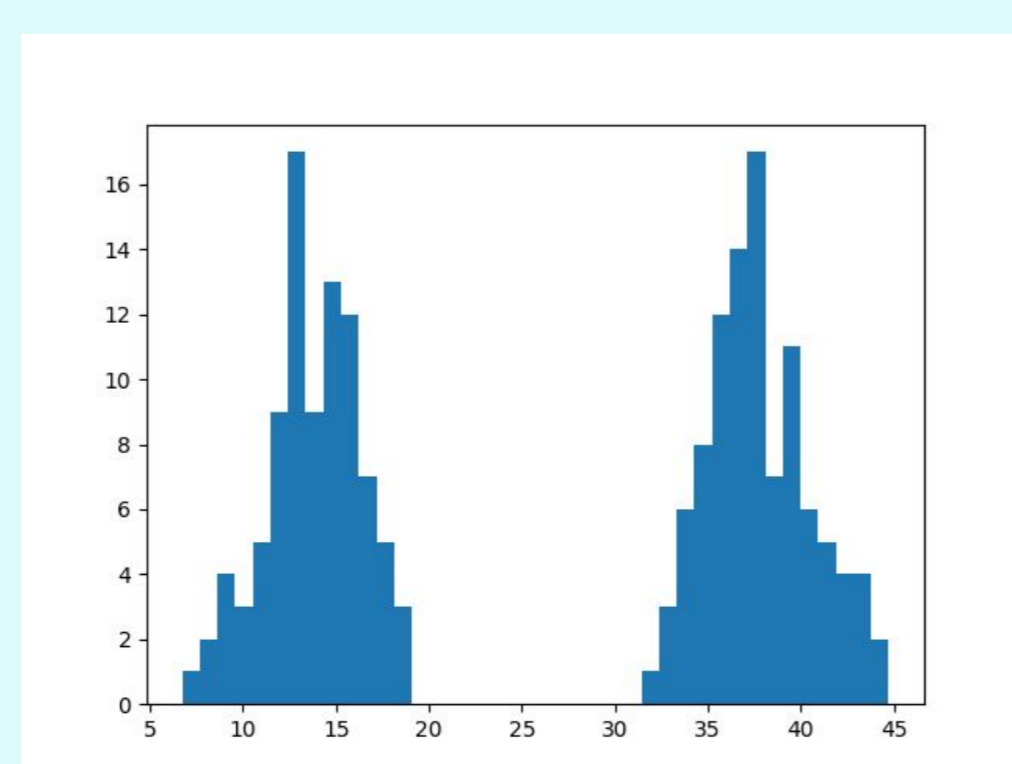
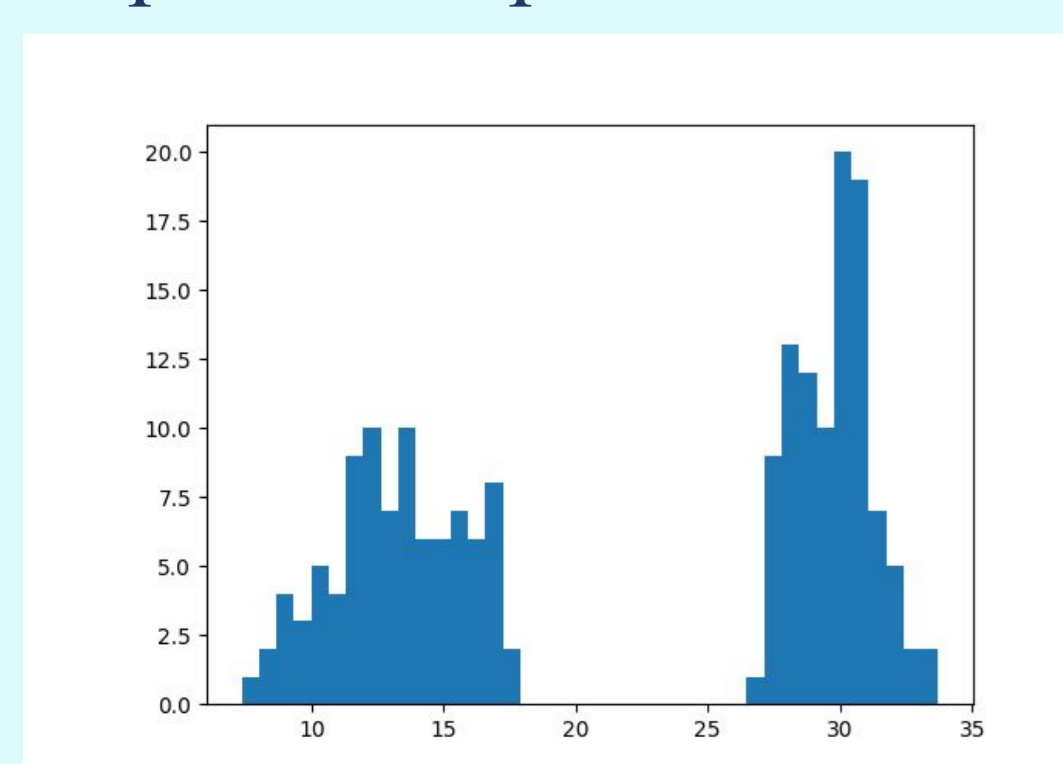
Methodology

In order to achieve our objectives, we decided to use the same molecules, that were used in research on RMSD, so our work can be more comparable to one done by scientists from University of North Carolina. Than we had to create a set of decoy structures, which would later be used in comparison with the original molecule. To make those we have used RNAsubopt tool from ViennaRNA package, which gave us a set of 2D predictions. Then we took received structures and the original ones, retrieved from PDB site, and loaded them into RNAComposer tool, which is used in prediction of 3D RNA structures. As a result of this, we have received a set of 3D configuration, which we could compare with each other. In order to do this, we have used previously mentioned MCQ4Structures tool.

Results

At this moment we have created histograms for distances between our decoy structures. Then we tested our score with normality tests, but in opposition to our assumptions, the data appeared to be not normally distributed. What’s even more interesting, for some molecules distribution appeared to be bimodal. We will have to take a closer look at this case.

We still have much more to do in this matter. Our research is far from the end, but it’s going in the right direction and we hope to present final answer to presented problem.



Workflow

Obtaining 2D structures of original molecules

Creating a set of 2D decoy structures

Creating a 3D models

Measuring distances between structures

Creating a histogram of given distances and comparing it to the normal distribution

GGAGUUCACCGAGGCCACGCGGAG
UACGAUCGAGGGUACAGUGAAUU

RNApdbec:

..(((((((...(((((((.....)))..)))..))))))..))))))

RNAsubopt:

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

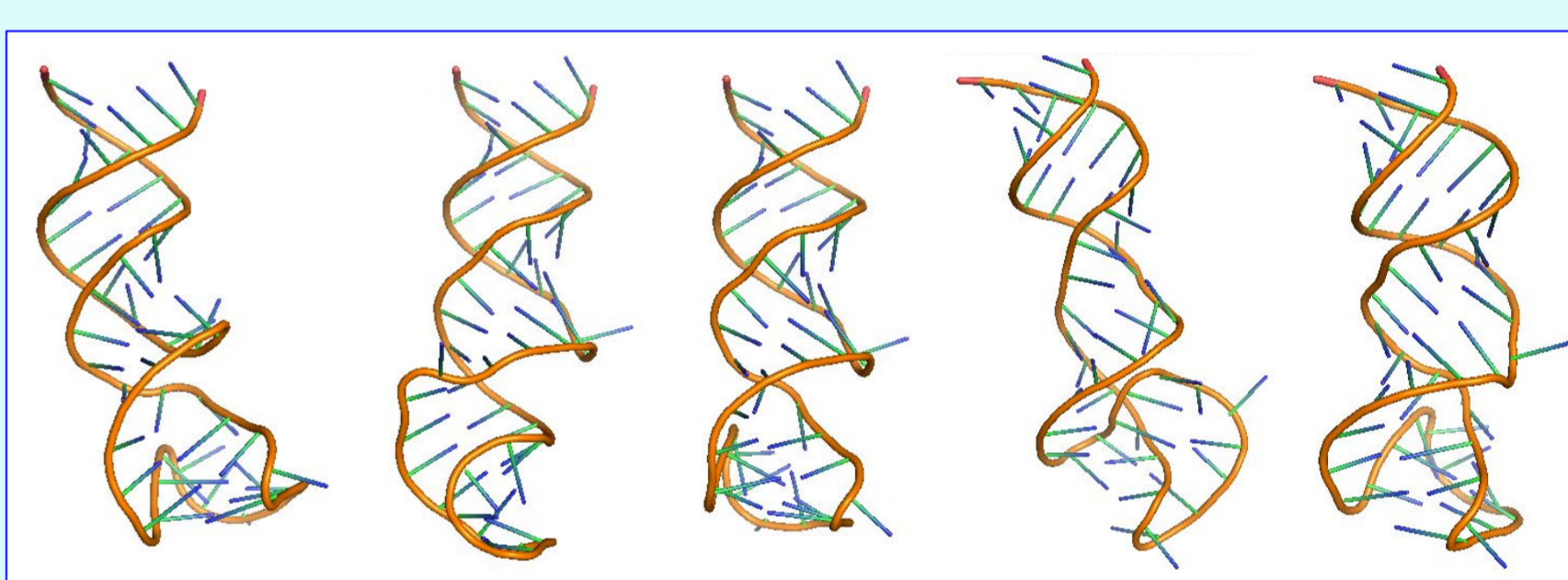
..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

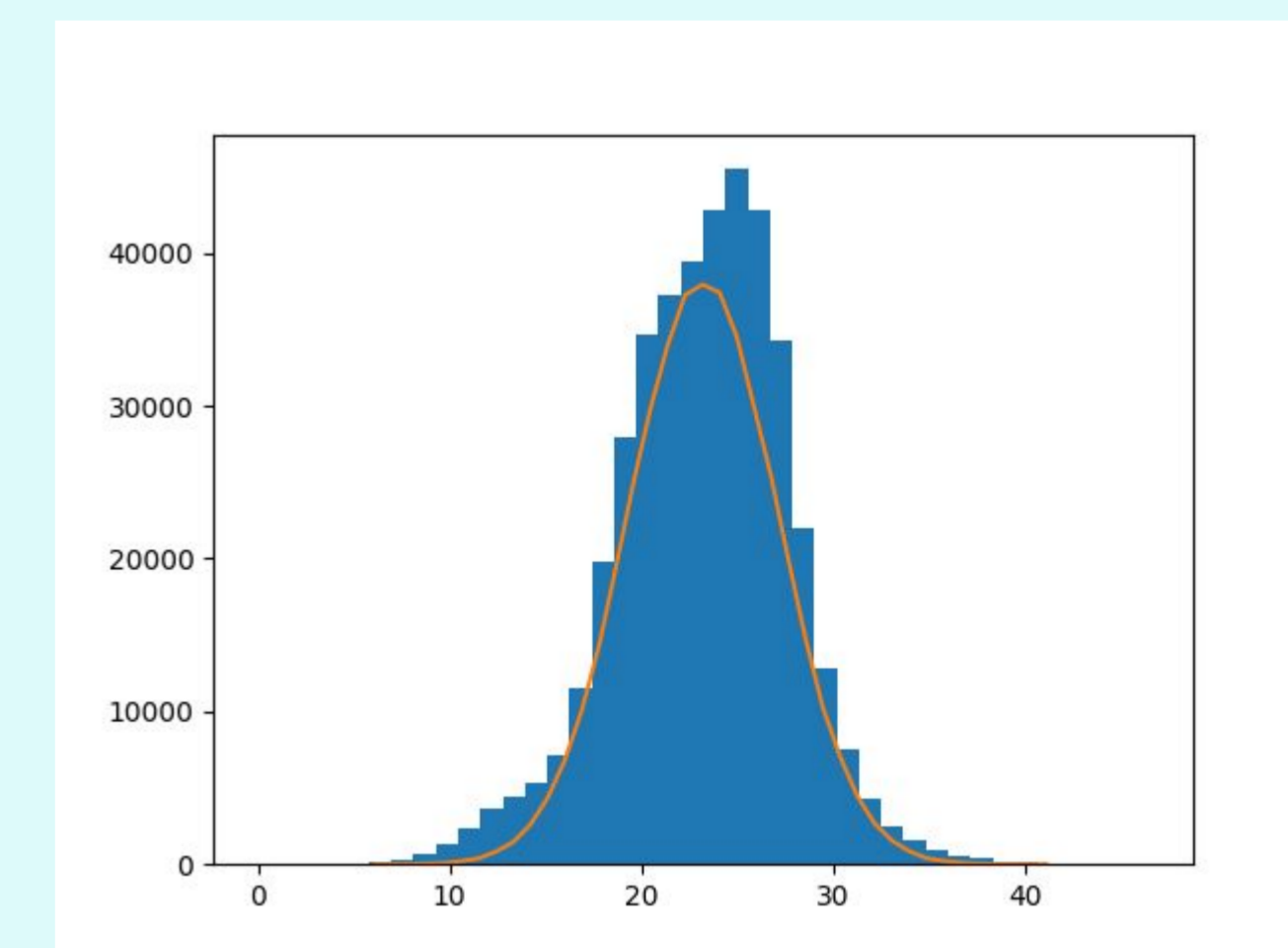
..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))

..(((((((...(((((((.....)))..)))..))))))..))))))



13	1XJR_subopt1_1	21.17825724883379	22.0215367833697	21.19205987001956	20.178
14	1XJR_subopt1_2	20.285790544053135	21.378189029510363	20.227132470704397	19.489
15	1XJR_subopt1_3	19.958034707850434	21.51726594809743	20.657010082418	19.951
16	1XJR_subopt1_4	20.608053134322468	21.58369299344332	20.76758044626844	20.544
17	1XJR_subopt1_5	21.754612138332728	22.639471515034714	21.899645588963665	21.955
18	1XJR_subopt1_6	19.47368269405073	21.515487150978128	20.303748856495712	19.765
19	1XJR_subopt1_7	20.562551584740863	21.285180993736148	21.170325317029157	20.917
20	1XJR_subopt1_8	20.90369783609336	22.06117832390136	20.920245866605626	20.986



References

- ^[1] Hajdin C, Ding F, Dokholyan N, Weeks K. (2010). On the significance of an RNA tertiary structure prediction. RNA (New York, N.Y.). 16. 1340-9. 10.1261/rna.1837410.
- ^[2] Zok T, Popena M, Szachniuk M. (2013). MCQ4Structures to compute similarity of molecule structures. Central European Journal of Operations Research. 22. 10.1007/s10100-013-0296-5.