

Structural bioinformatics

DrawTetrado to create layer diagrams of G4 structures

Michał Zurkowski¹, Tomasz Zok ¹ and Marta Szachniuk ^{1,2,*}

¹Institute of Computing Science, Poznan University of Technology, 60-965 Poznan, Poland and ²Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland

*To whom correspondence should be addressed.

Associate Editor: Yann Ponty

Received on February 7, 2022; revised on May 13, 2022; editorial decision on June 10, 2022; accepted on June 13, 2022

Abstract

Motivation: Quadruplexes are specific 3D structures found in nucleic acids. Due to the exceptional properties of these motifs, their exploration with the general-purpose bioinformatics methods can be problematic or insufficient. The same applies to visualizing their structure. A hand-drawn layer diagram is the most common way to represent the quadruplex anatomy. No molecular visualization software generates such a structural model based on atomic coordinates.

Results: DrawTetrado is an open-source Python program for automated visualization targeting the structures of quadruplexes and G4-helices. It generates static layer diagrams that represent structural data in a pseudo-3D perspective. The possibility to set color schemes, nucleotide labels, inter-element distances or angle of view allows for easy customization of the output drawing.

Availability and implementation: The program is available under the MIT license at <https://github.com/RNAPolis/drawtetrado>.

Contact: mszachniuk@cs.put.poznan.pl

1 Introduction

Quadruplexes are multilayered motifs occurring in nucleic acid structures. They tend to fold in guanine-rich regions, hence their abbreviated name G4. Every layer forms when four nucleotides arrange on a tetragonal plane and each of them makes pairings with two adjacent ones. Such layout of nucleotides is called a tetrad. Quadruplexes have a multitude of properties described by structural parameters. They include sequence, G-tracts, secondary structure topology, base-pair classification, nucleoside conformations, the number of stacked tetrads, tetrad planarity deviations, rise, twist, right- and left-handedness, torsion angles, the number of nucleic acids strands, strand polarity, type and length of loops, type and position of metal ions, etc. (Jana *et al.*, 2021; Zok *et al.*, 2022).

The complexity and specificity of the quadruplex are easier to understand if we have an appropriate visual model of its structure. However, not all models developed for nucleic acids are equally well suited to represent G4. Therefore, to visualize the secondary structure of tetrad or quadruplex, we introduced a dedicated top-down arc diagram, a two-line dot-bracket and a modified VARNA diagram (c.f. Fig. 1A)—basic versions of these representations could not reflect all base pairs that make up G4 motifs (Darty *et al.*, 2009; Popena *et al.*, 2020). The 3D structure of G4 can be shown using either of the existing visual models. The visualization type most often used in presentations and scientific publications is a layer diagram (c.f. Fig. 1B). To our knowledge, it cannot be automatically generated by any molecular visualization software. It presents a simplified model of a quadruplex highlighting its selected features (e.g. the number of tetrads, nucleoside conformations, the course of the

strand, the presence and types of loops). So far, the only visual model designed for the 3D structure of quadruplexes is cartoon-block schematics (Fig. 1C). These models are generated by DSSR-PyMOL integration and presented as static images of the structure viewed from six perspectives (Lu, 2020).

Here, we present DrawTetrado—an application to create layer diagrams of quadruplexes in DNA and RNA structures. They show the tetrads as a stack, each having four nucleobases colored according to anti or syn conformation. Strand directions are marked with arrows that support a visual identification of individual strands and determination of loop types (lateral, diagonal, propeller, V-shaped). The program automatically optimizes the model layout to give a readable image, even for complex cases like V-loops and G4-helices (quadruplex dimers). It allows customizing the diagrams and saving them in publication-quality SVG files. DrawTetrado is freely available at the GitHub repository.

2 Materials and methods

The DrawTetrado algorithm operates on the following data: G4-helix components, quadruplex components (the number of tetrads), G-tract components, tetrad types according to ONZ classification and nucleotide descriptions (type and conformation). These data are determined from the input 3D structure by the automatically run functions derived from ElTetrado (Zok *et al.*, 2020) and BpNet (Roy and Bhattacharyya, 2022). Then, DrawTetrado creates a layer diagram drawing in a multi-step procedure. At first, the algorithm determines the orientation of each tetrad. Then, it calculates

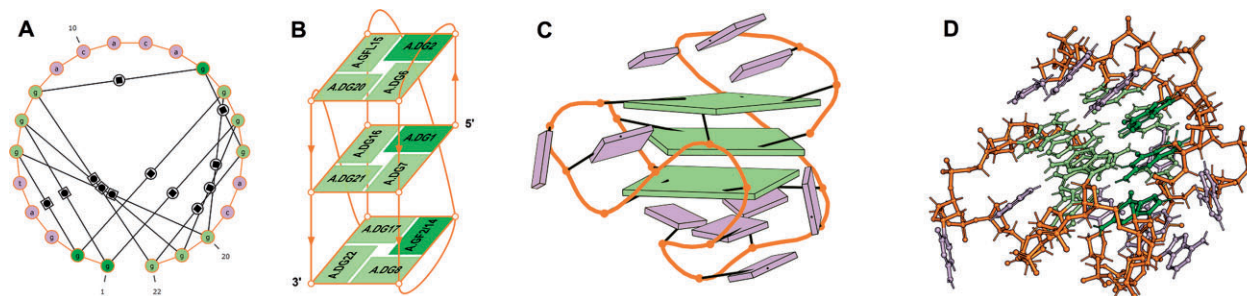


Fig. 1. Example visualizations of quadruplex structure, PDB ID: 6TCG (Haase and Weisz, 2020): (A) secondary structure diagram, (B) layer diagram from DrawTetrado, (C) cartoon-block schematics and (D) balls-and-sticks model

and draws the inter-tetrad connections located at the back and on the left-hand side of the diagram. In the next step, the algorithm connects nucleotides from the same layer. On top of this, it superimposes the shapes of nucleotides (parallelograms). Next, it frames the tetrads and draws the remaining inter-tetrad connections (the front and right-hand side ones). Finally, it labels all nucleotides in the diagram.

Connections between the tetrads are approximated by Bezier curves determined from the position of connected points and the curve orientation. The latter follows the polymer chain direction. Each connection is drawn separately. The algorithm distinguishes several types of links depending on the course and position of the curve. It tries to optimize the drawing for readability. Therefore, it prioritizes short vertical connections and applies penalties for the diagonal ones, especially those that run on the front of the diagram. The optimization takes place in the first step of the procedure when the rotation of tetrads is computed.

3 Using DrawTetrado

DrawTetrado works on all operating systems. It is written in Python 3.6+ and utilizes four extra modules—`pycairo`, `svgwrite`, `orjson` and `eltetrado`. The latter ones are automatically downloaded from the module repository while DrawTetrado installation. The internal optimization routine, implemented in C++, requires Cython and a C++20-compliant compiler. As the input, the program processes PDB and PDBx/mmCIF files. It can also accept JSON files generated by ElTetrado (Zok et al., 2020)—these contain quadruplex structural metadata determined from atomic coordinates, including contact network computed by the BPNet algorithm (Roy and Bhattacharyya, 2022).

The program is run via CLI (Command Line Interface) with one mandatory parameter—input file path—and many optional ones. It outputs layer diagrams in Scalable Vector Graphics (SVG) files. One can further edit them in any vector graphics software without quality loss.

Users can customize the drawing by modifying several parameters in the configuration file. They include the size (side lengths) of the nucleotide-representing parallelogram, the conformation-dependent color of the nucleotide (syn, anti, unrecognized), nucleotide label (label composition, font—typeface, color, size), spacing between nucleotides in the tetrad, the distance between layers (tetrads), the color of the tetrad frame, chain color and the viewing angle. The nucleotide label may consist of a chain identifier, a nucleotide name (short or long) and a nucleotide number. Changes to the configuration file are optional. The default parameters have been optimized to make the drawing readable and colorblind-friendly.

4 Conclusion

Bioinformatics resources are essential for studying biological data. So far, the quadruplex-dedicated ones have mainly focused on collecting and processing PQS (putative quadruplex sequences) (Miskiewicz et al., 2021). A few computational tools target 2D and 3D structures of G4s, including one for the 3D structure visualization (Lu, 2020). DrawTetrado responds to a growing demand for automated visualization of quadruplex structures in the most popular form—a layer diagram. It complements the collection of G4-dedicated tools created by the RNApolis team (Szachniuk, 2019). Since mid-2021, it has worked as a component of the ONQUADRO system (Zok et al., 2022) to visualize experimental PDB-derived quadruplex structures. Until now, it has created visualizations for 36 G4-helices and 599 quadruplexes stored in this database (data as of February 4, 2022). Available as a standalone program, it enables the creation of diagrams for arbitrary experimental and *in silico* G4 models.

Funding

We acknowledge support from Poznan University of Technology (statutory funds), and the National Science Centre, Poland [2019/35/B/ST6/03074].

Conflict of Interest: none declared.

References

- Darty, K. et al. (2009) Varna: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25, 1974–1975.
- Haase, L. and Weisz, K. (2020) Switching the type of V-loop in sugar-modified G-quadruplexes through altered fluorine interactions. *Chem. Commun. (Camb.)*, 56, 4539–4542.
- Jana, J. et al. (2021) Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.*, 2, 338–353.
- Lu, X.-J. (2020) DSSR-enabled innovative schematics of 3D nucleic acid structures with PyMOL. *Nucleic Acids Res.*, 48, e74.
- Miskiewicz, J. et al. (2021) How bioinformatics resources work with G4 RNAs. *Brief Bioinform.*, 22, bbaa201.
- Popenda, M. et al. (2020) Topology-based classification of tetrads and quadruplex structures. *Bioinformatics*, 36, 1129–1134.
- Roy, P. and Bhattacharyya, D. (2022) Contact networks in RNA: a structural bioinformatics study with a new tool. *J. Comput. Aided Mol. Des.*, 36, 131–140.
- Szachniuk, M. (2019) RNApolis: computational platform for RNA structure analysis. *FCDS*, 44, 241–257.
- Zok, T. et al. (2020) ElTetrado: a tool for identification and classification of tetrads and quadruplexes. *BMC Bioinformatics*, 21, 40.
- Zok, T. et al. (2022) ONQUADRO: a database of experimentally determined quadruplex structures. *Nucleic Acids Res.*, 50, D253–D258.