Untangling or further entangling? Revelation of the complicated world of nucleic acid quadruplex folds

Thenmalarchelvi Rathinavelan[*†] and Sruthi Sundaresan[*]

Department of Biotechnology, Indian Institute of Technology Hyderabad
Kandi, Telangana – 502284
[*]Equal contribution
[†]For correspondence: tr@bt.iith.ac.in

Keywords: G-quartets, Quadruplex fold, Non-G-quartets, Quadruplex motifs, Quadruplex modeling

Abstract
Quadruplexes are four-stranded nucleic acid secondary structures playing diverse biological roles. A recent meticulous analysis of known quadruplex structures formed from synthetic oligonucleotides shows that quadruplexes can take 33 distinct structural motifs (pair of stacked quartets) that act as scaffolds for simple (multi-layered quartets) and complex (multi-motifs) quadruplex folds. These motifs differ from one another in terms of interstrand connectivity and orientation. The connectivity is primarily governed by diagonal loops that connect alternate strands, and lateral and propeller loops that connect adjacent strands. These loops are absent in the tetramer, wherein interstrand orientation alone dictates the motif type. Nevertheless, a key question that remains unresolved is the complete structural landscape of quadruplex motifs, limiting our understanding of the fold-dependency in various biological functions and diseases. To address this, a total of 156 unique theoretically possible quadruplex motifs are mathematically devised here. A tool named 3D-NuS-Qplex (https://project.iith.ac.in/3d-nus-qplex/) has also been developed to model all the theoretically possible energy-minimized quadruplex folds with the options to incorporate overhangs and non-G-quartets (in addition to G-quartets) in both DNA and RNA contexts. The diversity of quadruplex motifs reported here provides new insight into their structural complexity, which appears greater than previously anticipated.

Introduction

Nucleic acids take a wide range of secondary structures besides the right-handed double helical B-form [1-5]. One among them is a four-stranded nucleic acid structure known as a quadruplex or tetraplex, which was discovered even before the discovery of the double helical structure [6]. There are several evidences about the biological role of these secondary structures formed by DNA and RNA sequences [7-16]. Besides, these structures also play a crucial role in several disease conditions [17].

About 500 quadruplex structures, mostly corresponding to synthetic oligonucleotides, have been solved predominantly using X-ray crystallography followed by NMR [18, 19]. These structures are formed by G-rich sequences, which form G-quartets, wherein four guanines interact with each other through Hoogsteen hydrogen bonds (**Figure 1A**, **1B**). Besides, nine non-G-quartets are also found to co-exist along with the G-quartets in a quadruplex structure (**Figure 2**).

A recent article has shown, based on the analysis of experimentally determined structures, that quadruplexes can be represented in a modular fashion wherein a pair of stacked quartets, namely, motif (**Figure 1C**), serves as the scaffold for quadruplex folds. These structural motifs can either be a monomer (formed by a single sequence), dimer (2 sequences), trimer (3 sequences), or tetramer (4 sequences). It is noteworthy that each of these motifs has 4 strands (refer to **Figure 1C** for the definition of strand numbers). The same study identified 33 different motifs through the analysis of quadruplex structures deposited in PDB, which differ among themselves by the relative orientations between the four strands and the loops that connect them [18, 19]. Notably, if a loop connects the adjacent strands that interact through base…base hydrogen bonds in a parallel and antiparallel orientation, it is called a propeller and lateral loop, respectively (**Figure 1D**). Further, the diagonal loop connects two alternate strands (which are not involved in base…base hydrogen bonding interaction) in an antiparallel orientation (**Figure 1D**). Although a loop can theoretically connect two alternate strands (not involved in base…base hydrogen bonding interaction) in a parallel orientation (**Figure 1D,** henceforth referred to as diagonal-parallel), it has not yet been observed in the experimental structures. Another point to note is that a motif can even be discontinuous, *viz.*, the adjacent quartets of all four strands in a quadruplex fold need not be connected [18]. A quadruplex fold can be formed using a single motif (**Figure 1E**), and a higher-order structure can be formed by combinations of motifs [18]. This further gives a clue that many undiscovered quadruplex structures may exist in nature, arising out of the combination of different motifs.

There are several attempts to develop algorithms (including machine learning) to predict the quadruplex folds from a given sequence [20-23]. Nonetheless, predicting the quadruplex structures from a given sequence remains a major challenge due to the lack of precise understanding of interstrand strand orientation and connectivity, which contribute to the diversity among quadruplex motifs. Additionally, structural diversity comes from a wide range of combinations that these motifs can adopt. This study explores all theoretically possible continuous monomeric, dimeric, trimeric, and tetrameric quadruplex motifs. Further, a tool, namely, 3D-NuS-Qplex, has been developed to model all the theoretically derived quadruplex folds.

Implementation of a web interface for the automated modeling of quadruplex folds

The 200 motifs, among which 156 are unique, can be extended to a quadruplex fold, *viz.,* a multi-layered G-quartet. This necessitates the need for the development of a tool that can model all the theoretically feasible quadruplex motifs to explore their biological relevance. Further,

only a limited number of motifs are reported in experimental structures [18]. Indeed, an existing web tool models only a limited number of quadruplex folds [24]. Here, a web server, namely 3D-NuS-Qplex (https://project.iith.ac.in/3d-nus-qplex/), is developed to model these quadruplex folds by applying a helical twist (30°) and rise (3.4 Å) between the adjacent quartets. Thus, to model all the theoretically possible quadruplex folds, 30° is chosen here as the helical twist. Similarly, the helical rise is chosen to be 3.4 Å. However, the values mentioned above are chosen to be constant across all the folds to model a broad range of quadruplex folds defined by all the motifs. Furthermore, 3D-NuS-Qplex also facilitates the use of nine different non-G-quartets (**Figure 2**), which are found alongside the G-quartets in experimentally determined structures deposited in the PDB [18, 19]. 3D-NuS-Qplex can also model quadruplex structures with 5' and/or 3' overhangs. The web interface of 3D-NuS-Qplex is shown in **Figure 3**.

Through the HTML interface (**Figure 3**), the user has to provide information about the type of quadruplex motif (by selecting the appropriate image of the motif) which is to be used as a template for the fold, sequence type (DNA or RNA), number of quartets in the quadruplex fold, input residues individually for the 4 quadruplex strands (representing the G-quartets as well as non-G-quartets, if any), loop sequence as applicable, 5' and 3' overhang residues for each sequence (if applicable) and position of the non-G residues as in each strand. Note that in the case of a monomer, the sequence type should either be DNA or RNA, since the four strands of the quadruplex are connected. However, for the dimer (4 combinations), trimer (8 combinations), and tetramer (16 combinations), the sequence can be either DNA, RNA, or a combination of both. Soon after retrieving the information, the HTML interface passes it to a PHP script. The PHP subsequently decodes the quadruplex motif type into a 10-letter alphanumeric pattern. For instance, 1a-2p|3a-4 represents an antiparallel dimer motif named DA7. Here, the first two strands (indicated by '1' and '2') are oriented in an antiparallel fashion (indicated by 'a'), second and third (indicated by '3') are oriented in a parallel fashion (indicated by 'p') and the third and fourth (indicated by '4') strands are oriented in an antiparallel fashion (indicated by 'a'). While '|' indicates the separation of two independent sequences, '-' indicates the two connected strands. Thus, the fold 1a-2p|3a-4 represents a dimer, wherein the strands 1 and 2 are formed by one sequence, and another sequence forms strands 3 and 4. Similarly, 1p-4p|3p-2 is a dimer (DP5) with a different connectivity. 1p-2p-3p-4 encodes for a monomer (named as MP1), wherein all four strands are connected in a parallel fashion. 1a|2a|3a|4 represents a tetrameric motif, QA3, in which the strands are in an antiparallel fashion with respect to each other. PHP passes this information to a bash script, which uses a FORTRAN code to generate the quadruplex fold coordinates and arranges the residue numbers as per the connectivity. Soon after the job is submitted, the user is provided with a molecule ID to retrieve the results later (**Figure 3**).

Quadruplex fold generation and optimization

As mentioned above, using the appropriate helical twist and rise, the quadruplex folds having G and non-G quartets are generated. For the propeller loop, lateral loop, diagonal loop, diagonal-parallel loop, 5'-overhang, and 3'-overhang, the representative coordinates are initially generated as per the nucleotide sequence by applying helical rise using the bash script. Thus, the bash script generates the initial coordinates for the quadruplex folds, with steric hindrances and a disjointed backbone. Although the diagonal-parallel loop (*viz.*, two alternate strands are connected in a parallel orientation) is not yet observed in any of the structures reported experimentally, the tool also models motifs having diagonal-parallel loops, since it is theoretically possible. Yet another point is that two crossing diagonals, which are present on the same side (DA3 motif), are also not observed experimentally but can be modeled by 3D-

NuS-Qplex. In these cases, the user must consider a longer loop sequence to avoid steric hindrances. Subsequently, the generated models are subjected to the Powel minimization algorithm using Xplor-NIH [25], an updated version of Xplor [26], to establish backbone connectivity and remove steric hindrance.

Initially, 1000 cycles of Powell minimization are carried out by constraining (*viz.*, not allowing them to move) the base atoms, leaving the backbone atoms free. During this round, only the dihedral energy terms are considered. For the subsequent rounds of minimization, quartet planarity and hydrogen bonding distance (using soft asymptotic potential) restraints, along with chi and sugar pucker restraints, are introduced. All the bonded energy terms are included, while van der Waals and electrostatic energy terms are included as discussed below. In the next step, the loop residues are kept free to remove the steric hindrance and establish the connectivity between adjacent nucleotides. If more than one loop is present (note that monomer has the highest number of 3 loops and tetramer has 0 loops), the loops are minimized one at a time. During this minimization cycle, all the quadruplex strands are also fixed. The minimization of a single loop occurs in six rounds, where each round involves 1,000 cycles of Powell minimization. During the first four rounds of minimization, atom-based nonbonded interactions (between the atoms whose mass is greater than zero) were considered with a cut-off value of 100 and by excluding the nonbonded interactions between bonded atoms, atoms that are bonded to a common third atom, and the atoms that are connected to each other through three bonds. While the electrostatic energy term is set to be off, the weightage of the van der Waals energy term is increased in each round (*viz.*, 0.002, 0.005, 0.01, and 0.5). At the fifth round, the electrostatic energy is included, and the van der Waals energy term is given a weightage of 1. Atom-based nonbonded interactions are considered with a cut-off value of 4.5, and the non-bonded interactions are excluded as mentioned above, but with the inclusion of 1-4 nonbonded interactions. During the last round, 1000 cycles for Powell minimization are extended by increasing the weightage of the dihedral restraints to 2. Similarly, overhang residues are also minimized. Soon after all the loop residues are individually minimized (while constraining other loop and quadruplex strands), all the loop residues are left free for the next round of minimization. During the next round, each strand of the quadruplex is minimized individually to remove steric hindrance and establish internucleotide connectivity following the abovementioned six rounds of minimization protocol. Here, initially, the last (fourth) strand is left free while constraining the first three strands. Subsequently, the 3rd, 2nd, and 1st strands are minimized one at a time. Following this, the fourth strand is minimized again as mentioned above, and with the inclusion of the 3rd, 2nd, and 1st strands in subsequent steps. In the final round, all strands, loops, and overhang residues are not constrained and are only restrained; six rounds of minimization are carried out as mentioned above.

The user can access the results using the molecule ID provided during submission. Finally, the user can interactively view the modeled quadruplex fold with JSmol enabled within the HTML interface (**Figure 3**). The user can also download the PDB and XPLOR input files. A PHP script controls the entire process. **Figure 3** illustrates the overall workflow of 3D-NuS-Qplex, and **Figures 4** and **5** provide representative examples of quadruplex models generated by 3D-NuS-Qplex.

Limitations

The models generated can only be considered as an initial model and not a representation of the global minima. Further, complex quadruplex folds arising out of the combinations of different motifs (involving both left-handed and right-handed motifs) are not possible in the current implementation.

Conclusions

A web tool, 3D-NuS-Qplex, is also developed here to model all the theoretically possible quadruplex folds, incorporating 5' and/or 3'-overhangs, non-G-quartets, and different combinations of DNA and/or RNA strands. Since the quadruplex structures are found to play a crucial role in several biological functions and diseases, this study brings a new dimension to the (i) understanding of quadruplex fold dependency on biological functions and diseases, (ii) fold-dependent design of ligands, and (iii) custom design of synthetic quadruplex folds for nanotechnological applications.

Author contribution

SS developed the theoretically possible quadruplex motifs, designed the HTML interface, and carried out testing. TR developed the methodology, implemented the modeling in PHP and bash scripting. SS and TR wrote the manuscript. TR designed and supervised the project.

Acknowledgements

The authors thank Dr Abhishek Kumar for the initial discussion on the 3D-NuS-Qplex web tool and the IITH computer center for technical support. SS thanks the Ministry of Education (MoE) for the fellowship.

Financial support

We greatly acknowledge the workstation support from ANRF (CRG/2022/001825).

Conflict of interest

The authors declare no conflict of interest.

References

1. Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995). Triplex DNA structures. Annu Rev Biochem. **64**, 65-95. https://doi.org/10.1146/annurev.bi.64.070195.000433.
2. Bacolla, A. & Wells, R. D. (2004). Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem. **279**, 47411-47414. https://doi.org/10.1074/jbc.R400028200.
3. Ravichandran, S., Subramani, V. K. & Kim, K. K. (2019). Z-DNA in the genome: from structure to disease. Biophys Rev. **11**, 383-387. https://doi.org/10.1007/s12551-019-00534-1.
4. Neidle, S. (2009). The structures of quadruplex nucleic acids and their drug complexes. Curr Opin Struct Biol. **19**, 239-250. https://doi.org/10.1016/j.sbi.2009.04.001.
5. Sathyaseelan, C., Vijayakumar, V. & Rathinavelan, T. (2021). CD-NuSS: A Web Server for the Automated Secondary Structural Characterization of the Nucleic Acids from Circular Dichroism Spectra Using Extreme Gradient Boosting Decision-Tree, Neural Network and Kohonen Algorithms. J Mol Biol. **433**, 166629. https://doi.org/10.1016/j.jmb.2020.08.014.
6. Bang, I. (1910). Untersuchungen über die Guanylsäure. Biochem. Z. **26**, 293-311.
7. Abiri, A., Lavigne, M., Rezaei, M., Nikzad, S., Zare, P., Mergny, J. L. & Rahimi, H. R. (2021). Unlocking G-Quadruplexes as Antiviral Targets. Pharmacol Rev. **73**, 897-923. https://doi.org/10.1124/pharmrev.120.000230.
8. Saranathan, N. & Vivekanandan, P. (2019). G-Quadruplexes: More Than Just a Kink in Microbial Genomes. Trends Microbiol. **27**, 148-163. https://doi.org/10.1016/j.tim.2018.08.011.
9. Cammas, A. & Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res. **45**, 1584-1595. https://doi.org/10.1093/nar/gkw1280.

10. Tateishi-Karimata, H. & Sugimoto, N. (2021). Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. Nucleic Acids Res. **49**, 7839-7855. https://doi.org/10.1093/nar/gkab580.
11. Miglietta, G., Russo, M. & Capranico, G. (2020). G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. Nucleic Acids Res. **48**, 11942-11957. https://doi.org/10.1093/nar/gkaa944.
12. Lyu, K., Chow, E. Y., Mou, X., Chan, T. F. & Kwok, C. K. (2021). RNA G-quadruplexes (rG4s): genomics and biological functions. Nucleic Acids Res. **49**, 5426-5450. https://doi.org/10.1093/nar/gkab187.
13. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. Nat Rev Mol Cell Biol. **21**, 459-474. https://doi.org/10.1038/s41580-020-0236-x.
14. Spiegel, J., Adhikari, S. & Balasubramanian, S. (2020). The Structure and Function of DNA G-Quadruplexes. Trends Chem. **2**, 123-136. https://doi.org/10.1016/j.trechm.2019.07.002.
15. Griffin, B. D. & Bass, H. W. (2018). Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function. Plant Sci. **269**, 143-147. https://doi.org/10.1016/j.plantsci.2018.01.011.
16. Rhodes, D. & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. **43**, 8627-8637. https://doi.org/10.1093/nar/gkv862.
17. Ajjugal, Y., Kolimi, N. & Rathinavelan, T. (2021). Secondary structural choice of DNA and RNA associated with CGG/CCG trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. Sci Rep. **11**, 8163. https://doi.org/10.1038/s41598-021-87097-y.
18. Sundaresan, S., Uttamrao, P. P., Kovuri, P. & Rathinavelan, T. (2024). Entangled World of DNA Quadruplex Folds. ACS Omega. **9**, 38696-38709. https://doi.org/10.1021/acsomega.4c04579.
19. Uttamrao, P. P., Sundaresan, S. & Rathinavelan, T. (2023). Structure and Folding Patterns of RNA G-Quadruplexes. In RNA Structure and Function (Barciszewski J ed), pp. 205-232, Springer International Publishing, Cham.
20. Liew, D., Lim, Z. W. & Yong, E. H. (2024). Machine learning-based prediction of DNA G-quadruplex folding topology with G4ShapePredictor. Sci Rep. **14**, 24238. https://doi.org/10.1038/s41598-024-74826-2.
21. Cagirici, H. B., Budak, H. & Sen, T. Z. (2022). G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. BMC Bioinformatics. **23**, 240. https://doi.org/10.1186/s12859-022-04782-z.
22. Sahakyan, A. B., Chambers, V. S., Marsico, G., Santner, T., Di Antonio, M. & Balasubramanian, S. (2017). Machine learning model for sequence-driven DNA G-quadruplex formation. Sci Rep. **7**, 14535. https://doi.org/10.1038/s41598-017-14017-4.
23. Jurkowski, M., Kogut, M., Olewniczak, M., Glinko, J. & Czub, J. (2025). Large-Scale Conformational Analysis Explains G-Quadruplex Topological Landscape. J Phys Chem B. **129**, 9622-9632. https://doi.org/10.1021/acs.jpcb.5c04372.
24. Patro, L. P. P., Kumar, A., Kolimi, N. & Rathinavelan, T. (2017). 3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-Dimensional Nucleic Acid Structures. J Mol Biol. **429**, 2438-2448. https://doi.org/10.1016/j.jmb.2017.06.013.
25. Bermejo, G. A., Tjandra, N., Clore, G. M. & Schwieters, C. D. (2024). Xplor-NIH: Better parameters and protocols for NMR protein structure determination. Protein Sci. **33**, e4922. https://doi.org/10.1002/pro.4922.
26. Brünger, A. T. (1993). X-PLOR Manual.

Main text figures

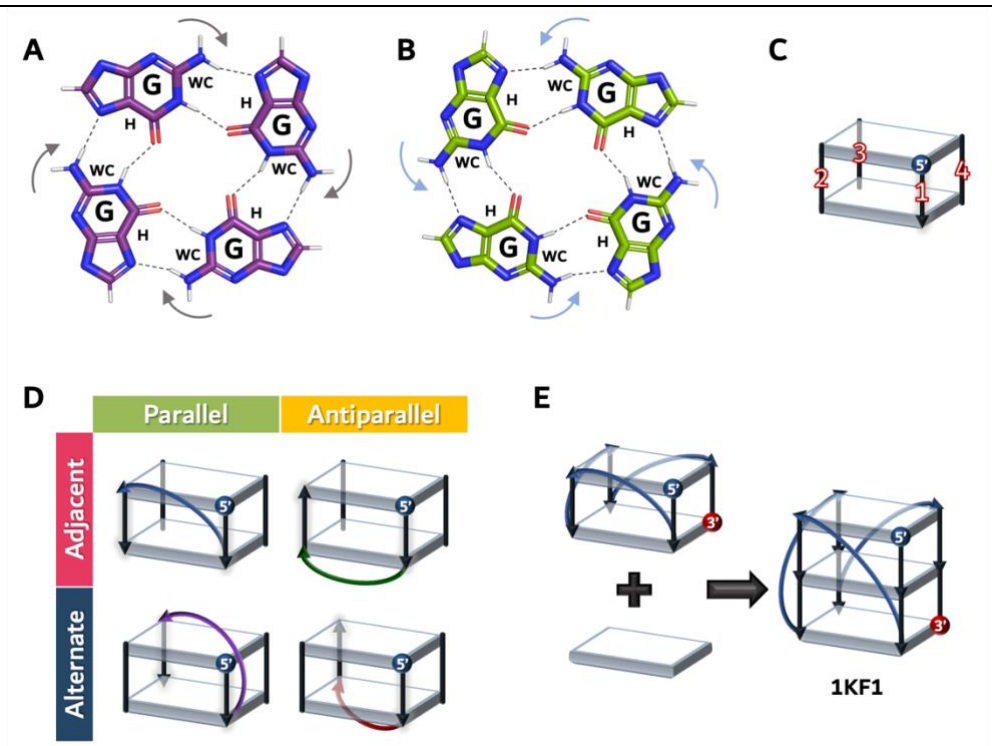

Figure 1. A schematic illustration of A) clockwise and B) anticlockwise Hoogsteen hydrogen bonding pattern (indicated by arrows) in a G-quartet. Hydrogen bonds are indicated in the dotted lines. "WC" and "H" represent Watson-Crick and Hoogsteen edges respectively. Note that the orientation of WC and H edges dictates the hydrogen bonding pattern. For example, when the WC edge of the bottom-right G is on the left side, it leads to the formation of a clockwise quartet (Left in A). One the other hand, if it is on the right side, leads to an anticlockwise quartet (Right in B). C) A representative illustration of a two layered quartet motif which acts as a building block for a fold. The number nomenclature marked here represents the strand numbers (1 to 4) of a quadruplex motif/fold and is followed throughout this article. The first strand always starts from the front-right with the 5'-end of the strand at the top. Note that the 5' (blue colored circle) and 3' (arrow head) ends of the 1st strand alone is indicated for the illustration purpose since the directionality of the other strands varies depending on the motif type [18]. D) The definition of propeller (colored blue), lateral (colored green), diagonal-parallel (colored purple) and diagonal (colored red) loops connecting the different quadruplex strands are shown. E) An example illustrating how a motif can lead to the formation of a simple quadruplex fold [18].

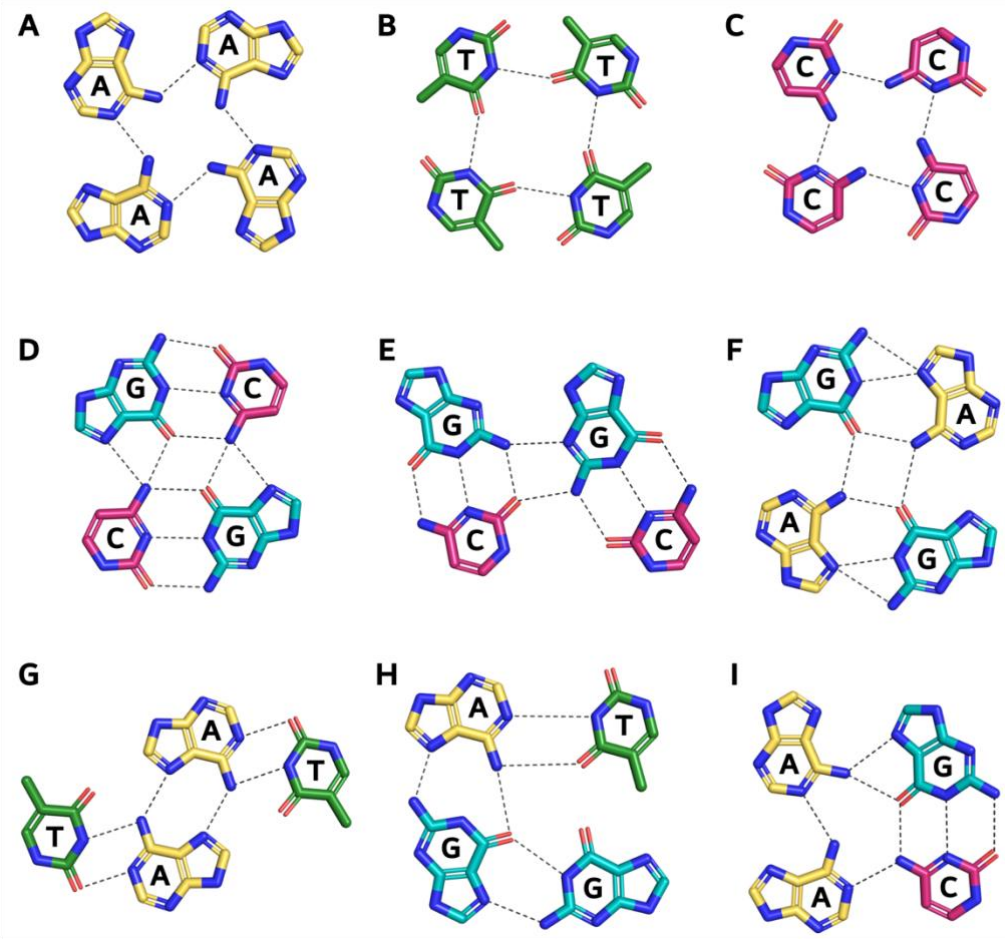

Figure 2. Different non-G-quartets: A) A-A-A-A, B) T-T-T-T, C) C-C-C-C, D) G-C-G-C, E) G-G-C-C, F) G-A-G-A, G) T-A-T-A, H) A-T-G-G and I) A-G-C-A. Dotted lines indicate hydrogen bonds.

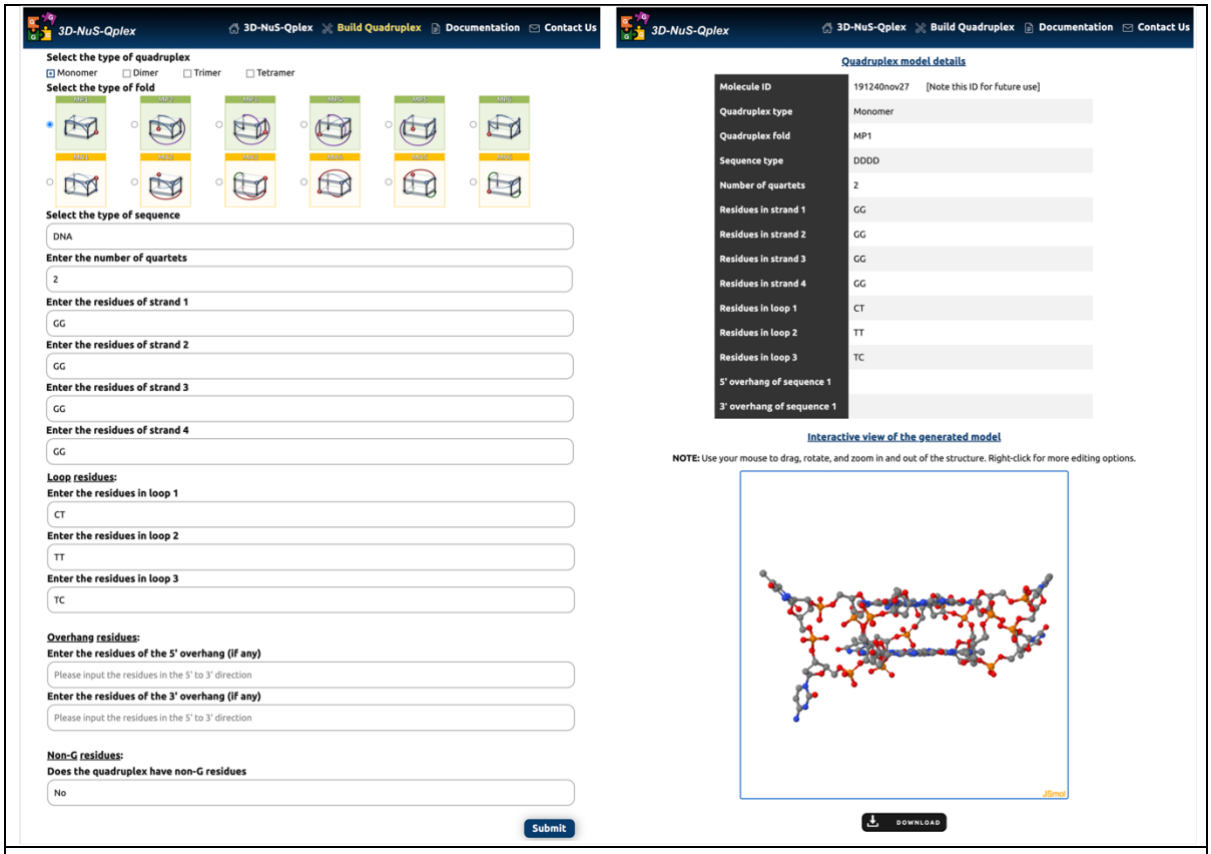

Figure 3: Snapshots of the 3D-NuS-Qplex interface. The workflow begins with the input interface shown on the left, where the user has to select the quadruplex type, fold type, the type of sequence (DNA/RNA), number of quartets, residues in each strand, loops, overhang residues (if any), and the positions of the non-G quartets (if any). After submitting the required inputs to model the quadruplex structure, the output interface on the right displays a consolidated summary of these inputs, and an interactive view of the generated model using JSmol. A video tutorial that succinctly explains the usage of the tool can be found at: https://youtu.be/U0VhdGoxP_I.

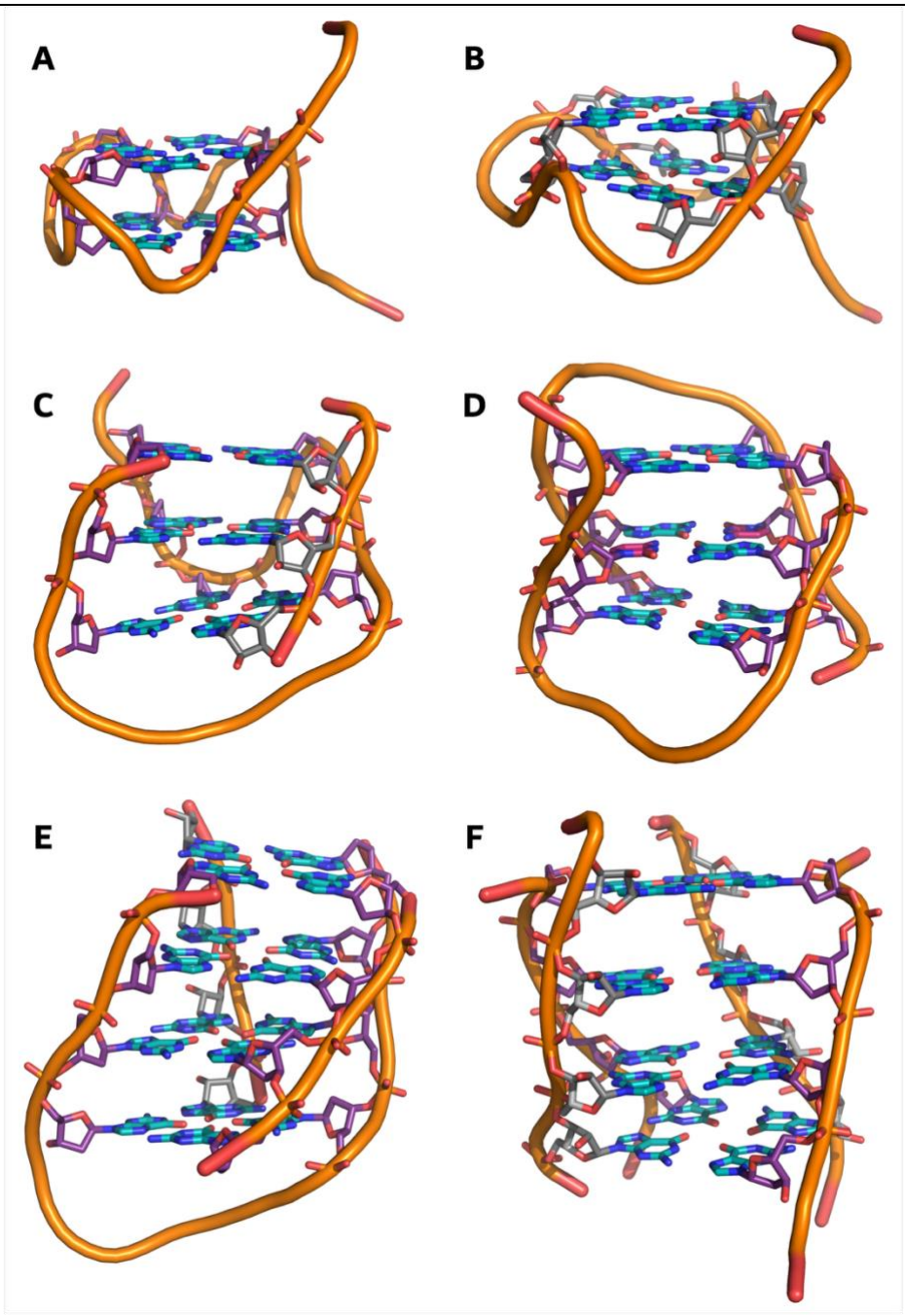

Figure 4. Models generated by 3D-NuS-Qplex: A) Monomeric DNA structure of MP1 with 5' and 3' overhangs, B) Monomeric RNA equivalent structure of MP1 with 5' and 3' overhangs, C) Dimeric structure of DA20 with the first sequence as RNA and the second sequence as DNA having a 5' overhang in the first sequence and a 3' overhang in the second sequence, D) Dimeric structure of DA14 having a 3' overhang in the first sequence and a 5' overhang in the second sequence along with a non-G quartet (G-C-G-C) quartet in the middle, E) Trimeric structure of TA3 composed of 3 sequences (DNA-RNA-DNA), and F) Tetrameric structure of QH1 composed of 4 sequences (DNA-RNA-DNA-RNA). Note that guanine residues are in blue, cytosine residues are in pink, deoxyribose sugars in purple, and ribose sugars is in grey.

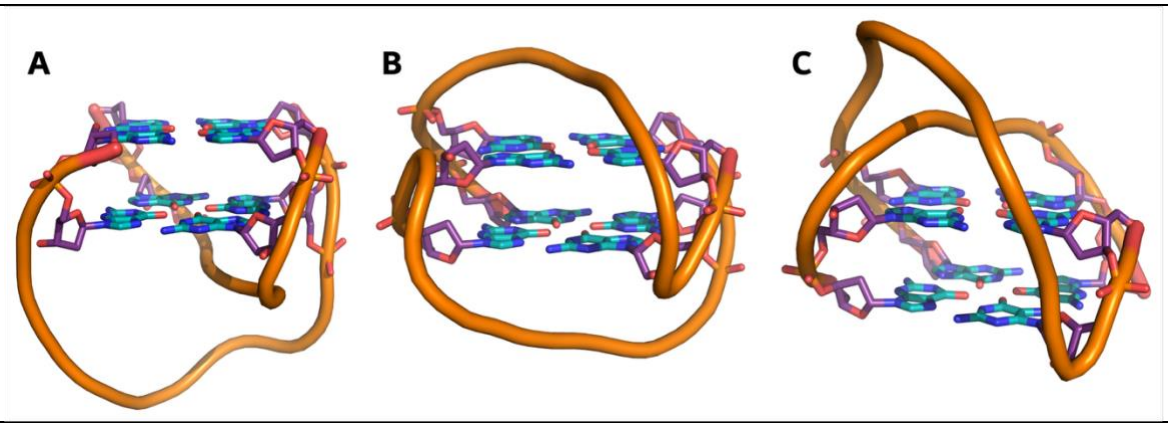

Figure 5. Quadruplex structures with crossing diagonal loops generated by 3D-NuS-Qplex: Dimeric structure A) DA3 with two diagonal loops that cross over each other, B) DA15 with two diagonal parallel loops, and C) DH3 with one diagonal loop and one diagonal-parallel loop that cross over each other. Note that guanine residues are in blue, the deoxyribose sugars in purple, and ribose sugars is in grey.