



Regular Article

Prediction of the secondary structure of short DNA aptamers

Arina Afanasyeva¹, Chioko Nagao² and Kenji Mizuguchi^{1,2}

¹Artificial Intelligence Center for Health and Biomedical Research (ArCHER), National Institutes of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka 567-0085, Japan

²Laboratory of In-silico Drug Design, Center for Drug Design Research (CDDR), National Institutes of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka 567-0085, Japan

Received July 1, 2019; accepted July 29, 2019

Aptamers have a spectrum of applications in biotechnology and drug design, because of the relative simplicity of experimental protocols and advantages of stability and specificity associated with their structural properties. However, to understand the structure-function relationships of aptamers, robust structure modeling tools are necessary. Several such tools have been developed and extensively tested, although most of them target various forms of biological RNA. In this study, we tested the performance of three tools in application to DNA aptamers, since DNA aptamers are the focus of many studies, particularly in drug discovery. We demonstrated that in most cases, the secondary structure of DNA can be reconstructed with acceptable accuracy by at least one of the three tools tested (Mfold, RNAfold, and CentroidFold), although the G-quadruplex motif found in many of the DNA aptamer structures complicates the prediction, as well as the pseudoknot interaction. This problem should be addressed more carefully to improve prediction accuracy.

Key words: G-quadruplex, dot-bracket notation, Tanimoto similarity

Corresponding author: Arina Afanasyeva, Bioinformatics project, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan.
e-mail: arina@nibiohn.go.jp

Biological drugs such as monoclonal antibodies have brought in a new era in medicine, providing novel treatments for diseases that are difficult to treat with small-molecule medicines. However, these drugs have also resulted in escalation of medical costs and decrease in patient quality of life. It is a pressing challenge, therefore, to find potentially drug-gable sites on the protein surface so that antibody drugs could be converted to small-molecule ones. To achieve such an ambitious goal, an efficient tool is required for "probing" protein surfaces and identifying key interactions.

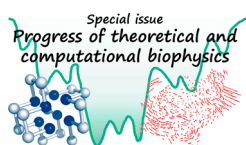
Aptamers are a relatively short oligonucleotide or peptide molecule that can bind selectively to target molecules of different types, such as proteins or small-size chemicals. Usually, nucleic acid (NA) aptamers adopt some particular globular structure, which determines its nuclease stability as well as increased selectivity because of the surface complementarity to a target molecule [1–3].

DNA oligonucleotide molecules, in general, are more stable than RNA, due to the lack of a hydroxyl group in DNA sugars, which make them less reactive. Structural differences, such as differences in helix form, also make DNA molecules more stable to the nucleases [4]. For these reasons, DNA aptamers are often more practical for use in biomedical approaches and studies.

DNA aptamers have promising applications in inhibiting protein-protein interactions (PPIs). Successful examples

◀ Significance ▶

DNA aptamers have many applications in biotechnology. In particular, they demonstrated the great ability as sensors of the protein surface, which can be helpful for the design of highly selective inhibitors of protein-protein interactions. The structure of the protein-aptamer complex is necessary for this kind of analysis, although most of the nucleic acid structure modelling programs are originally designed for RNA. In this study, we present the first attempt to assess the accuracy of three secondary structure predicting tools in application to DNA aptamers.



include the development of highly selective inhibitors of alpha-Thrombin [5] and Interleukin-6 (IL6) [6] and a study of the conformational properties of Human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) [7].

In addition, in comparison to antibodies, DNA aptamers can be more easily modified to modulate binding affinity to a target molecule by regulating the intermolecular contact surface. Information on the affinity changes, therefore, can offer an insight into the important interactions on the PPI surface, which can be used further to design small molecule inhibitors with desirable properties. Such an attempt has great implications for establishing a next-generation drug discovery platform; DNA aptamers can be utilized as sensors for finding potentially druggable sites on the protein surface to convert antibody drugs to small-molecule ones.

In the development of this drug discovery platform, determining the aptamer-target complex structure is an essential element. However, experimentally solving the structure, typically by X-ray crystallography, is time-consuming and may be difficult for some target proteins. Computational modeling of DNA aptamer-protein complex structure would be an attractive alternative approach. It should consist of several steps, including secondary structure prediction, followed by the 3D structure reconstruction of the aptamer and sampling the aptamer/protein complex structure by docking or other appropriate methods of protein-DNA complex prediction.

In this study, we discuss the first step of the pipeline—the reconstruction of the secondary structure of DNA aptamer from the sequence.

Secondary structure describes which nucleotides form Watson-Crick base pairs and which are located in loops. In many cases, additional interactions may be formed between loop nucleotides or a free strand and the stem parts, as in the three-stranded type of folding, when nucleotides of a free DNA/RNA strand are intercalated between the base pairs of double-stranded DNA/RNA. Such interactions are called pseudoknots. A special case of non-Watson-type interaction is G-quadruplex structural formation. Such formation increases thermodynamic stability and stabilizes the aptamer structure [5]. Hence, G-quadruplex is often found in DNA aptamers.

Several computational programs/algorithms already exist for the prediction of NA structure. However, most of these tools were designed for RNA [8–11], because of the availability of a variety of biological single-stranded (ss) RNA molecules and RNA aptamers. While most of these programs can be superficially applied to DNA, at least for secondary structure prediction, no such comprehensive analysis has been published so far to assess the capability of these tools in predicting the secondary structure of DNA aptamers. Our current study aims to present the first assessment of this kind.

In this study, we assessed three main approaches to predict the 2D structure of RNA/DNA. Mfold developed by Zuker in 1980s [12] has historically been the first approach based on finding the structure with the minimal free energy (MFE approach) making use of the energy parameters

obtained from the thermodynamic experiments [13]. The following modifications of Mfold and the other MFE-based approaches such as RNAfold [10] have included the partition function implementation for assessing the base pair probabilities, which have shown considerable improvement in prediction accuracies [10]. Later a different approach arose. In contrast to Mfold, based on the searching of the optimal or suboptimal folding of the RNA sequence, the new approach was focused on the analysis of the ensemble of all possible solutions with the centroid estimator. One example of this approach implementation is CentroidFold utilizing the generalized centroid estimator or γ -estimator [11].

We present the prediction accuracy of the Mfold, RNAfold, and CentroidFold in application to short-length DNA aptamers. Furthermore, we provide an insight into the G-quadruplex motif types found in crystal structures of DNA aptamers, since aptamers with G-quadruplex are particularly difficult for modeling due to G-quadruplex diversity and complexity of possible motives.

Methods

Evaluation of the NA secondary structure prediction methods on the test dataset of the short-length DNA aptamers with available crystal structures

1. Construction of the test dataset

We started our analysis from preparing the test set of the DNA aptamers with available crystal structures. For that purpose, we searched all entries in Protein Data Bank (PDB) [14] containing DNA.

1. For the state of the database on the January 30th, 2019, we downloaded 6546 PDB entries of single DNAs and DNA in complex with proteins and other molecules.
2. From these structures, we extracted only DNA/RNA molecules (some entries contained both DNA and RNA), excluding all the remaining molecules such as proteins, small molecules, ions, and waters.
3. The extracted DNA/RNA structures were automatically annotated using the tool *analyse* from the 3DNA software suite [15]. From the resulted annotation, we extracted the information on the number of chains and presence/absence of the protein in the original PDB structure, the number of nucleic acids, number, and type of base pairs and the presence of G-quadruplex.
4. Based on the information from the 3DNA annotation, we further extracted only single-stranded DNA (887 entries)
5. From the resulted set of ssDNA entries, we filtered out short chain structures with less than 25 NA and filtered out structures not forming any Watson-Crick or G-quadruplex interactions.

The resulted dataset contained 69 DNA structures with the length in the range of 25–57 nucleotides with the various

types of folding. Thirty-two aptamers were originally in complex with proteins; these examples might be useful for the further analysis of aptamer-protein complex prediction. Twenty-six aptamers contain G-quadruplex structural element, some of them are fully formed by G-quadruplex, while others have G-quadruplex as a part of the structure.

2. Evaluation of prediction accuracy of the 2D structure modeling methods

There are several types of representation of the DNA/RNA secondary structure including the graphical representation with several variations of 2D diagram types and text representations, for example, column text representation, where information on the paired bases is presented in two columns of residue numbers. One of the most commonly used formats is the dot-bracket representation. In the context of dot-bracket annotation, the whole chain is presented as a single string, where positions of the paired nucleotides are shown with matching parentheses and unpaired nucleotides with dots. Classic Watson-Crick base pairs are usually presented by round parenthesis, and pseudoknots could be indicated by square or curly brackets. We decided to use this format since it is the most commonly used in 2D structure prediction programs. Although within the scope of this study, we did not focus on the prediction of the pseudoknots. Positions of the G-quadruplex motif are assigned with '+' as set in RNAfold software.

To assess the accuracy of the 2D prediction, we used the single string format of the dot-bracket representation. This representation allows comparison of resulting annotations to the original by calculating the coefficient of similarity of two strings of the same length with the Tanimoto similarity score as follows:

$$\text{Score} = N_{\text{ident}}/N_{\text{sum}},$$

where N_{ident} - the number of matching positions in two strings; N - string length

Example:

- Original: ((((((((((.....))((.....))(((.....))))))))))
- Predicted: ((((((((((.....(((.....)))).....)).....))))))
- Score: 0.74

For the original crystal structures of aptamers in the test dataset, first, we obtained 2D annotations in the column text format with 3DNA software [15] and then converted into a dot-bracket format with the simple Bourne shell script.

3. Protocol for running the structure prediction programs

We evaluated three programs for 2D structure prediction: Mfold, Centroid_fold, and RNAfold. These three methods are freely available as standalone software programs. After installation, these programs can be run from the command line, which makes it convenient to run automatically for the

big set of data. These tools accept fasta-formatted sequence files as the only input.

RNAfold and CentroidFold only work with RNA sequences; therefore, we modified original sequences, replacing 'T' by 'U'. Non-natural residues were replaced by the corresponding natural nucleic acids.

Centroid_fold and RNAfold result in one predicted structure while Mfold can generate several predictions depending on the running parameters, such as folding temperature, ionic concentration of the Na^+ and Mg^{2+} , and others. Mfold algorithm results in one or several optimal and suboptimal structures for a given sequence based on the calculated energy, although eventually, crystal structure or bound structure of the aptamer is not necessarily the one with the lowest energy. We found that in many cases, the 'correct' structure was found within suboptimal solutions. For the test, we set the type of molecule (NA) to 'DNA' and varied the parameter of the percent suboptimality (P) from 5 to 20% (default is 5%) to generate at least ten possible solutions for each of the test aptamers. Other parameters were set to default values.

RNAfold was installed as a part of the Vienna RNA program suit [10]. This program generates one minimal energy solution. To run the program, we used the option to calculate g-quadruplexes (-gquad) since many of the test set aptamers contain G-quadruplex structural element.

Centroid fold program prediction is based on the calculation of a base-pairing probability matrix for RNA sequence. The program makes use of several algorithms: the McCaskill, the CONTRAfold (default) and *pfold* model from the Vienna RNA package. To run the program, we used default settings for all parameters.

Results and discussion

To date, many different measures for comparison of DNA/RNA secondary structures in dot-bracket format have been developed, including base pair distance, which counts the number of different base pairs in two structures, the Hamming distance between two symbolic-notated sequences, the tree edit distance [16] based on tree representations of secondary structures and some other measures [17–19]. However, methods like RNAdistance [17], calculating base-pair distance, cannot handle G-quadruplex, and therefore can not be applied to the third part of our test set. For that reason, we chose Tanimoto score, since it can be applied to all instances in the test set and uniformly assess 2D structure prediction accuracy, although above mentioned scores could be useful to analyse stem-loops types of aptamers separately.

The resulting accuracy scores for three programs Mfold, RNAfold, and CentroidFold, are presented in summary table (Table 1), where '1' means 100% correct prediction, i.e., all of the paired/unpaired nucleotide positions are correctly defined, as well as all guanines involved in G-quadruplex formation. The accuracy range 0.8–0.9 usually indicates solution close to the correct one with one or several missing/excessive

Table 1 Summary table of the accuracy of the 2D structure prediction programs on the test set of DNA aptamers

CentroidFold	RNAfold	Mfold	PDB_ID	Structure	Len
1.	1.	1.	1JVE	Triplex-DNA	27
1.	1.	1.	1NGU	Hairpin with pseudoknots	27
1.	1.	1.	2OEY	Hairpin with loops	25
1.	1.	1.	2VWJ	Hairpin with dangling ends	26
1.	1.	1.	3THW	Hairpin with loops	53
1.	1.	1.	4HT4	Hairpin with dangling ends	28
0.92	1.	1.	1NGO	Hairpin	27
0.92	1.	1.	5N2Q	Hairpin with dangling ends	26
0.77	1.	1.	3H25	Hairpin with dangling ends	27
0.55	1.	1.	1AW4	Hairpin with pseudoknots	27
1.	0.96	1.	6CCE	Hairpin with dangling ends	57
0.6	0.6	1.	1B4Y	Triplex-DNA	30
0.91	0.95	0.95	2N8A	Double hairpin	45
0.75	0.65	0.95	3HXQ	Two-forked with pseudoknots	41
0.88	0.88	0.94	5HRT	Hairpin with loops	34
0.93	1.	0.93	1GN7	Triplex-DNA	32
0.93	1.	0.93	1WAN	Triplex-DNA	32
0.8	1.	0.93	2ARG	Hairpin with pseudoknots	30
0.93	0.93	0.93	4F41	Hairpin	32
0.93	0.93	0.93	4F43	Hairpin	32
0.93	0.93	0.93	4TMU	Hairpin with dangling ends	29
0.87	0.93	0.93	4ER8	Hairpin with dangling ends	32
0.75	0.75	0.93	5HRU	Hairpin with loops and pseudoknots	32
0.84	0.92	0.92	134D	Triplex-DNA	25
0.84	0.92	0.92	135D	Triplex-DNA	25
0.84	0.92	0.92	136D	Triplex-DNA	25
0.9	0.9	0.9	5D2Q	Hairpin	40
0.89	0.89	0.89	4CEI	Hairpin with dangling ends	37
0.83	0.89	0.89	4CEH	Hairpin with dangling ends	37
0.88	0.92	0.88	5LD2	Hairpin with dangling ends	51
0.83	0.88	0.88	3U44	Hairpin with dangling ends	36
0.61	0.66	0.88	1SNJ	Two forked	36
0.5	0.66	0.88	1EZN	Two forked	36
0.85	0.85	0.85	3U4Q	Hairpin with dangling ends	27
0.61	0.73	0.85	2F1Q	Three-forked	42
0.8	0.7	0.85	4REC	Double-stranded	40
0.92	0.84	0.84	1OMH	Hairpin with dangling ends	25
0.92	0.84	0.84	1QX0	Hairpin with dangling ends	25
0.92	0.84	0.84	1S6M	Hairpin with dangling ends	25
0.92	0.84	0.84	1ZM5	Hairpin with dangling ends	25
0.76	0.84	0.84	5D23	Hairpin	26
0.83	0.88	0.83	5D2S	Hairpin	36
0.33	0.73	0.73	2M91	Hairpin with G-quadruplex	30
0.8	0.45	0.72	3HXO	Two forked with pseudoknots	40
0.62	0.7	0.7	2M8Z	Hairpin with G-quadruplex	27
0.77	0.68	0.67	4CEJ	Hairpin with dangling ends	46
0.65	1.	0.63	5CMX	Double-stranded with G-quadruplex	30
0.62	0.87	0.62	2M90	Hairpin with G-quadruplex	32
0.14	0.2	0.61	2M92	Hairpin with G-quadruplex	34
0.5	0.56	0.56	2M93	Hairpin with G-quadruplex	32
0.48	0.48	0.51	4I7Y	Double-stranded with G-quadruplex	27
0.53	1.	0.46	2HY9	G-quadruplex	26
0.53	1.	0.46	2JPZ	G-quadruplex	26
0.53	1.	0.46	2LPW	G-quadruplex	26
0.53	1.	0.46	5MVB	G-quadruplex	26
0.53	1.	0.46	6CCW	G-quadruplex	26
0.52	1.	0.44	2JSL	G-quadruplex	25
0.52	1.	0.44	2JSQ	G-quadruplex	25
0.48	0.92	0.4	2MBJ	G-quadruplex	27
0.48	0.64	0.4	2M53	G-quadruplex	25
0.46	0.92	0.38	5Z80	G-quadruplex	26
0.58	0.38	0.38	5MTA	G-quadruplex	34
0.58	0.38	0.38	5MTG	G-quadruplex	34
0.42	0.71	0.35	2MS9	G-quadruplex	28
0.42	0.71	0.35	4U5M	G-quadruplex	28
0.42	0.71	0.32	2N3M	G-quadruplex	28
0.1	1.	0.25	201D	G-quadruplex	28
0.36	1.	0.24	5J6U	G-quadruplex	25
0.1	1.	0.14	230D	G-quadruplex	28

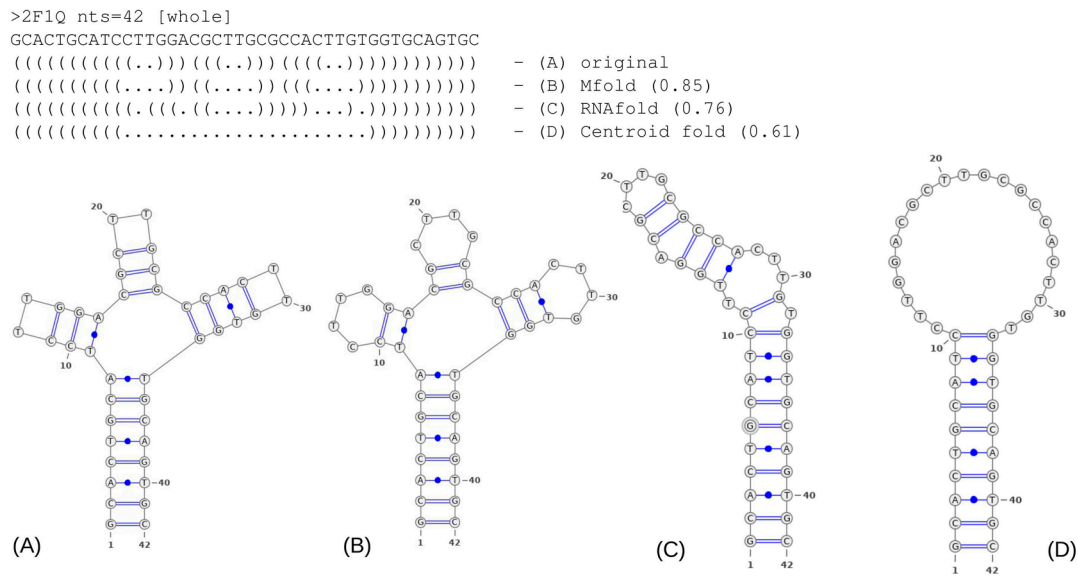


Figure 1 Example of the 2D structure predictions obtained by B) Mfold, C) RNAfold and D) Centroid fold in comparison to the original 2D structure A) of 2F1Q DNA aptamer. On the top dot-bracket representations, Tanimoto similarity scores for predicted topologies are shown in brackets.

base pairs but overall correctly determined stem/loop parts. It should be noted that for Mfold, we assessed the best-resulted solution, which was not necessarily the top scored, while for RNAfold and CentroidFold, it was a single top scored solution. RNAfold is the only method out of three capable of predicting G-quadruplex formation, and in most of the cases, the prediction was correct. In a few cases (PDB ID: 2MS9, 4U5M, 2N3M), RNAfold correctly found one of the tetrads of G-quadruplex but failed to define the second.

To provide a better understanding of the accuracy scores, we present in Figure 1 an example of predicted 2D structures vs. 2D structure of the crystal structure of the three-forked DNA aptamer (PDB ID: 2F1Q; diagram representation, obtained by VARNAs software with NView drawing algorithm [20]). The best prediction was generated by Mfold and had an accuracy of 0.85 (as estimated by the Tanimoto similarity coefficient as described in the Methods section). The best-predicted structure has three missing base pairs, although the overall folding is correct, therefore, this result can be acceptable in most cases, since we assume that the missing interactions may be restored further over the 3D structure reconstruction procedure, for example, by additional energy minimization and relaxation procedures.

In summary, for 26 out of 69 aptamers, the 100%-accurate structure was found with at least one of the algorithms, which accounts for 38% of the dataset. Mfold test resulted in 36 out of 69 (52%) aptamer structures predicted with acceptable accuracy of >0.85 ; RNAfold accurately predicted 44 structures (64%), which includes most of the G-quadruplex-containing structures; Centroid fold 25 aptamers (36%). Therefore, the best performance on the DNA aptamers test set was demonstrated by the RNAfold program. For the 14 aptamers (20%), all 3 algorithms failed to determine a solu-

tion with acceptable accuracy of >0.85 . Of these 14 difficult cases, 11 entries are presented by the aptamers with G-quadruplex, which indicates the complexity of G-quadruplex motif reconstruction.

Furthermore, we analyzed the efficiency of the scoring functions of the tested programs in distinguishing the accurate prediction (Fig. 2). This analysis is especially relevant for Mfold since this program generates several alternative solutions.

Interestingly, for all of the three tested programs, we found similar behavior of the scoring functions. We did not find clear correlations between calculated energies and accuracy of the predicted 2D structure (calculated as described in the Methods section). Nevertheless, from the scatter plots presented in Figure 2, it is clear that predictions with relatively high accuracy (>0.8) tend to have more widely spread scores while poorly predicted structures scored in the more narrow dispersion of energies close to 0.

Analysis of G-quadruplex folding types - investigation of the possible dependence of G-quadruplex folding on the NA sequence

G-quadruplex is the structural motif naturally formed in guanine-rich sequences in DNA or RNA. When four guanines laying in the same plane are connected to two neighbor guanines by two H-bonds forming so-called tetrads or G-tetrads, several G-tetrads stacked on top of each other form G-quadruplex. G-quadruplex is often stabilized by the metal ion positioned in the center of the folding such as Na^+ or K^+ . Naturally, G-quadruplexes are found at the telomeric regions of chromosomes where they act as protection of the telomere ends.

There are a variety of possible folding types for

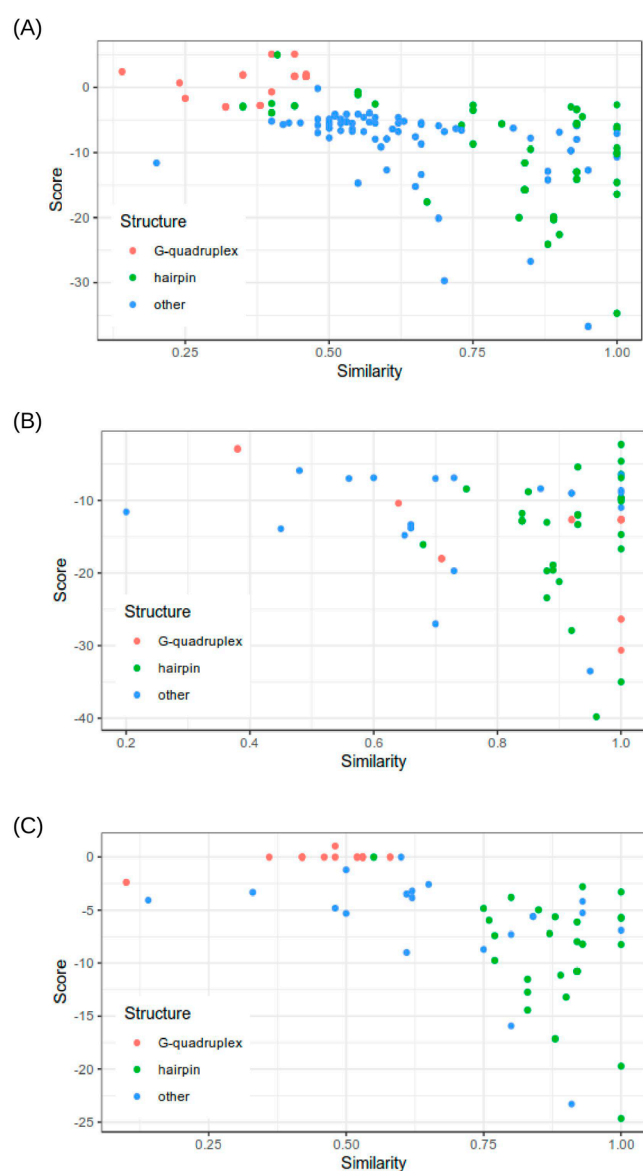


Figure 2 Distribution of the 2D structure prediction scores resulted by A) Mfold, B) RNAfold and C) Centroid fold depending on the accuracy of the predicted structures.

G-quadruplex, which considerably complicates the task of aptamer structure prediction, although there are several methods that can handle the problem of structure prediction for G-quadruplex containing aptamers. Such programs as RNAfold and also other methods are designed specifically for the prediction of G-quadruplex positions such as QGRS Mapper (Quadruplex forming G-Rich Sequences) [21], although these programs predict only positions of guanines, forming G-quadruplex, this information is not enough to assume the possible folding type for the further 3D structure modelling.

G-quadruplexes can be classified depending on the folding type as shown in [22] and in Figure 3, although in practice,

various foldings of G-quadruplex found in crystal structures do not comply with these types. In many cases observed G-quadruplex folding has some kind of hybrid folding type or modification of those types that usually described in papers [22–24]. We analyzed the PDB database of possible G-quadruplex foldings and tentatively divided them into several groups depending on the folding type, although this segregation is approximate and does not reflect any standard classifications (Fig. 4). The number of chains that form G-quadruplex can affect the resulted folding. Type 1 structure represents ‘tetramolecular parallel’ folding type (Fig. 3), type 6 corresponds to ‘unimolecular chair’ folding, and type 9 to ‘unimolecular basket,’ while other structures presented in Figure 4 shows some hybrid or more complicated foldings.

Analysing the resulted groups we have found that they include clusters of aptamers with conserved sequence motives, therefore, we performed multiple sequence alignment of such sequence clusters (Fig. 5 and Supplementary Figures S1–S3). We found that one structural group may include several sequence motives, and some motives can form two types of foldings (Fig. 5B). For example, aptamers with the sequence motif ‘GGTTGGTGTGGTTGG’ always forms the folding number 6 and sequence ‘TGAGGGTG GXGAGGGTGGGTAAGG’ forms type 3 folding. Although some sequences, for example, ‘TAGGGTTAGGGTTAGGG TTAGGG’ apparently can form several foldings (type 2 or type 9), possibly the folding in such cases is defined by some thermodynamic factors and interaction with other molecules. The groups with type 9 and type 2 foldings consist of the most variable sequences, while in other groups some particular sequence patterns can be traced. We speculate that the sequence of the G-quadruplex may affect the folding type since the length of the straights between G regions will restrict the length of the loops and introduce conformational limitations or additional stacking interactions.

These findings may be useful for the aptamer structure reconstruction. The tentative type of the G-quadruplex can be suggested based on the G-quadruplex sequence, and appropriate G-quadruplex structure, therefore, can be used as a template.

Conclusion

We demonstrated that DNA aptamer secondary structure could be reconstructed with reasonable accuracy by the commonly used tools, although there are several implications associated with DNA. We mainly concentrated on the G-quadruplex motifs since it is often introduced in DNA aptamers to improve nuclease stability, and also adds another layer of complexity to the problem of structure prediction. The analysis of G-quadruplex suggested that template-based modeling was possible for some types of G-quadruplex foldings. In this work, we did not address another critical issue of the pseudoknot prediction. The programs for pseudoknot

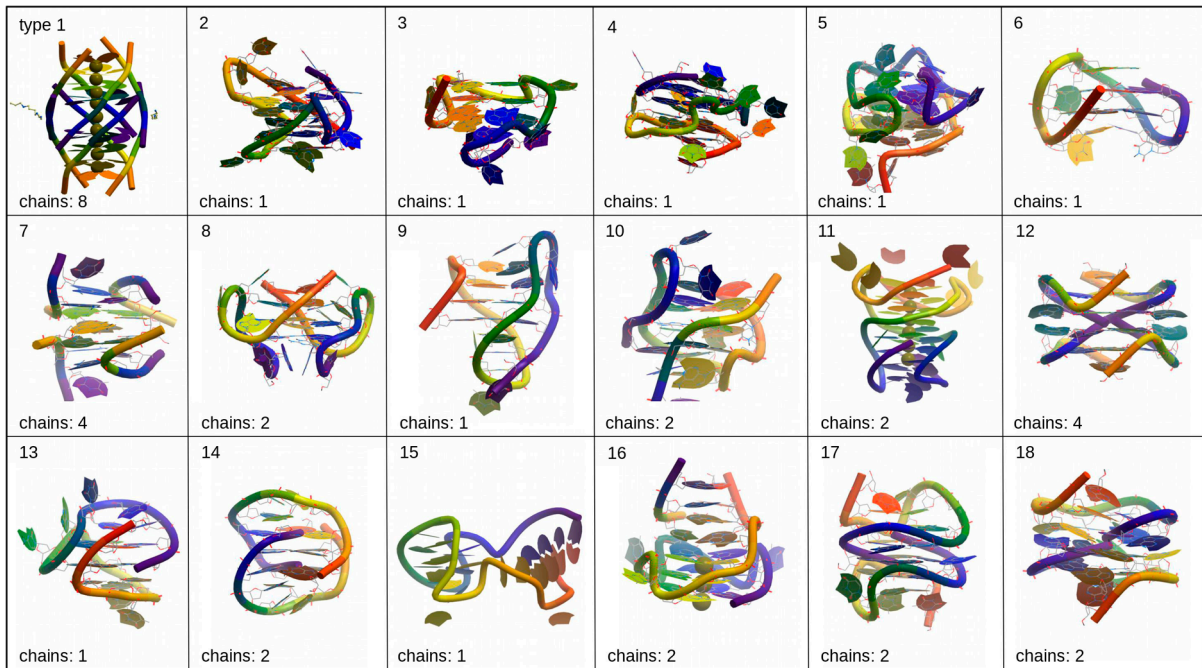


Figure 3 Examples of the various G-quadruplex folding types (PDB structures are shown in ribbon representation in rainbow color scheme with 5'-end colored purple and 3'-end – red).

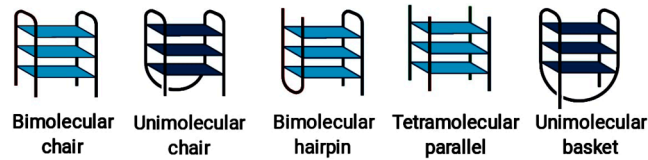


Figure 4 Schematic representations of the G-quadruplex common folding types.

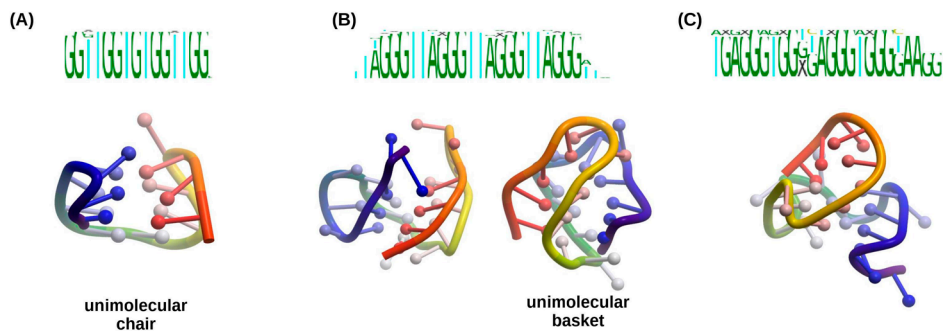


Figure 5 Examples of folding types corresponding to specific sequence motifs: A) type 6 folding, B) types 2 and 9 foldings are formed by the aptamers with same sequence motifs; C) type 3 folding.

prediction usually require higher computational cost [25,26], although pseudoknot interactions may be crucial for further 3D structure reconstruction. Our analysis of the accuracy of 2D structure prediction tools for DNA aptamers is expected to be useful for aptamer-based drug design.

Acknowledgment

This research was supported by AMED under grant number JP18ak0101102.

Author Contributions

A. A. designed and carried out analysis and wrote the manuscript. C. N. and K. M. managed the entire project and co-wrote the manuscript.

Conflict of Interest

A. A., C. N. and K. M. declare that they have no conflict of interest.

References

- [1] Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nat. Rev. Drug Discov.* **9**, 537–550 (2010).
- [2] Missailidis, S. & Hardy, A. Aptamers as inhibitors of target proteins. *Expert Opin. Ther. Pat.* **19**, 1073–1082 (2009).
- [3] Rohloff, J. C., Gelinias, A. D., Jarvis, T. C., Ochsner, U. A., Schneider, D. J., Gold, L., *et al.* Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids*. **3**, e201 (2014).
- [4] Kratschmer, C. & Levy, M. Effect of chemical modifications on aptamer stability in serum. *Nucleic Acid Ther.* **27**, 335–344 (2017).
- [5] Russo Krauss, I., Pica, A., Merlino, A., Mazzarella, L. & Sica, F. Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 2403–2411 (2013).
- [6] Gelinias, A. D., Davies, D. R., Edwards, T. E., Rohloff, J. C., Carter, J. D., Zhang, C., *et al.* Crystal structure of interleukin-6 in complex with a modified nucleic acid ligand. *J. Biol. Chem.* **289**, 8720–8734 (2014).
- [7] Das, K., Balzarini, J., Miller, M. T., Maguire, A. R., DeStefano, J. J. & Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for nucleotide incorporation vs pyrophosphorolysis-binding of foscarnet. *ACS Chem. Biol.* **11**, 2158–2164 (2016).
- [8] Gong, S., Wang, Y., Wang, Z. & Zhang, W. Computational methods for modeling aptamers and designing riboswitches. *Int. J. Mol. Sci.* **18**, E2442 (2017).
- [9] Andersen, E. S. Prediction and design of DNA and RNA structures. *N. Biotechnol.* **27**, 184–193 (2010).
- [10] Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
- [11] Sato, K., Hamada, M., Asai, K. & Mituyama, T. CentroidFold: a web server for RNA secondary structure prediction. *Nucleic Acids Res.* **37**, W277–W280 (2009).
- [12] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
- [13] SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **95**, 1460–1465 (1998).
- [14] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- [15] Lu, X.-J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
- [16] Shapiro, B. A. An algorithm for comparing multiple RNA secondary structures. *Comput. Appl. Biosci.* **4**, 387–393 (1988).
- [17] Lorenz, R., Bernhart, S. H., Höner zu Siederdisen, C., Tafer H., Flamm, C., Stadler, P. F., *et al.* ViennaRNA Package 2.0. *Algorithms. Mol. Biol.* **6**, 26 (2011).
- [18] Agius, P., Bennett, K. P. & Zuker, M. Comparing RNA secondary structures using a relaxed base-pair score. *RNA* **16**, 865–878 (2010).
- [19] Moulton, V., Zuker, M., Steel, M., Pointon, R. & Penny, D. Metrics on RNA secondary structures. *J. Comput. Biol.* **7**, 277–292 (2000).
- [20] Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
- [21] Kikin, O., D’Antonio, L. & Bagga, P. S. QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682 (2006).
- [22] Đapić, V., Abdomerović, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J. O., *et al.* Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.* **31**, 2097–2107 (2003).
- [23] Kwok, C. K. & Merrick, C. J. G-Quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* **35**, 997–1013 (2017).
- [24] Hao, F., Ma, Y. & Guan, Y. Effects of Central Loop Length and Metal Ions on the Thermal Stability of G-Quadruplexes. *Molecules* **24**, E1863 (2019).
- [25] Huang, X. & Ali, H. High sensitivity RNA pseudoknot prediction. *Nucleic Acids Res.* **35**, 656–663 (2007).
- [26] Seetin, M. G. & Mathews, D. H. TurboKnot: Rapid prediction of conserved RNA secondary structures including pseudoknots. *Bioinformatics* **28**, 792–798 (2012).

This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

