MJFAS MALAYSIAN JOURNAL OF FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

RESEARCH ARTICLE

Comparison of similar RNA 3D structures and substructures search tools

Margareta Deidre Valeska [a], Arli Aditya Parikesit [b,*]

[a] *Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences (i3L), East Jakarta 13210, DKI Jakarta, Indonesia*
[b] *Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences (i3L), East Jakarta 13210, DKI Jakarta, Indonesia*

* *Corresponding author: aril.parikesit@i3l.ac.id*

Article history
Received 2 September 2019
Revised 20 December 2019
Accepted 12 February 2020
Published Online 16 August 2020

Abstract

ncRNAs functions in the cells has produced exponentially increasing data of RNA 3D structures in both PDB and NDB. Tools are needed to search for similar RNA 3D structures in hope to aid with annotation processes. In this article, five tools (ARTS, FASTR3D, RNA FRABASE 2.0, RAG-3D, and R3D-BLAST2) will be compared based on their performances in analyzing samples of RNA 3D structures in different sizes (pseudoknot, 5S rRNA, and 18S rRNA). ARTS was found to be the most outdated and slow tools, while FASTR3D and RNA FRABASE 2.0 is commonly used as a benchmarking standard. RAG-3D & R3D-BLAST2 produces results with the highest accuracy and is still relevant in the present, however, RAG-3D is found to be limited to analyze RNA 3D structures to a certain size (it cannot process very large samples). This article also provides suggestion for future studies to create tools with similar purposes.

Keywords: RNA 3D structures, RAG-3D, R3D-BLAST2, FASTR3D, RNA FRABASE 2.0, ARTS

INTRODUCTION

RNAs that aren't translated into proteins, also known as non-coding RNAs (ncRNAs), has become an increasing topic in bioinformatics research in recent years due to their functional roles in cells such as mRNA modifications and gene regulation [1, 2]. Principally, the functions of RNAs are thought to be determined by the formation and folding of their 3D structures. This has caused an increased in the number of RNA information stored in the PDB [3] and NDB [4], especially in terms of RNA 3D structures. Annotation of RNA structures and functions have also become increasingly difficult due to the large amount of data being updated regularly. Because of this, tools and programs that are able to efficiently and effectively search the PDB and/or NDB databases for similar RNA 3D structures and substructures are in high demand to help with the annotation process.

Recent search tools such as RAG-3D [5], R3D-BLAST2 [6], ARTS [7, 8], FASTR3D [9], and RNA FRABASE 2.0 [10], use some form of automatic and heuristic approach to scan the PDB database for similar RNA 3D structures and/or substructures according to the user input. They also deploy an easy-to-use online user interface, where a user is only required to provide a PDB file or RNA sequence for the program to work. Both FASTR3D and RNA FRABASE 2.0 use a form of pattern-based searching algorithm based on the user's input to search the PDB database for similar secondary (2D) structures. Meanwhile, RAG-3D and R3D-BLAST2 searches the PDB database for similar tertiary (3D) structures according to the user's input using two different algorithmic approaches. RAG-3D uses a graph representation algorithm based on the graph topology of RNA 3D structures, while R3D-BLAST2 uses a structural-alphabet (SA) alignment algorithm that turns all 3D structures of RNA into 1D SA sequences.

The main goal of this review is to deliver an overview of how each tool works based on their algorithm and database creation, as well as how they compare to each other. This review article will also touch on several topics regarding how these tools have improved the research and understanding of RNA functions and its future applications.

EXPERIMENTAL

Methods

Four journal articles were collected from PubMed and Google Scholar. These journals were obtained on May 20, 2019 using "RNA 3D structures algorithms" as the keyword and between 2005–2018. All selected journals were to be of the most recent version and is still widely used due to their relevancy in the field. Each of these journals focuses on algorithms and interfaces for RNA 3D structure similarity based on secondary and tertiary structures. Two of them uses algorithms based on RNA secondary structures and the other two are based on RNA tertiary structures. Table 1 shows the review summary of each journal. Each algorithm will be discussed in this review article.

Table 1 Elemental analysis by XRF for Ag@TiO_2 microspheres.

Author	Tool Name	Algorithm	Database	Input Type	Input Size	Output/Result	Max Time Needed
Dror et al., 2005	ARTS	Two adjacent base pairs stored in hash table and compared	Non-redundant RNA structures from comparison	PDB file and email address	All RNA sizes	Table with PDB ID, no. of nucleotides & base pairs, RMSD, 3D alignment	Several minutes
Lai et al., 2009	FASTR3D	Hashing algorithm based on derived secondary structure	Derived secondary structures of RNA in hash tables	Manual input (1D, 2D, 3D sequences)	All RNA sizes	Table summary with PDB ID, chain ID, 1D-3D RNA sequence, class, method	Several seconds
Popenda et al., 2010	RNA FRABASE 2.0	Pattern-searching algorithm based on derived atom coordinates	Derived atom coordinates in different categories	Manual input (1D, 2D sequences)	All RNA sizes	Table with PDB & NDB ID, Chain ID, 1D-3D RNA sequence, class, method	Several seconds
Zahran et al., 2015	RAG-3D	3D graph representation based on topology	RNA 3D graphs & subgraphs (2-10 vertices)	PDB ID or PDB file	Small RNA (limited to vertices)	Query of secondary motifs & subgraphs, table of similar structures, 3D view	Several minutes
Yen et al., 2017	R3D-BLAST2	SA-based algorithm based on nucleotide cluster	1D SA sequences of RNA 3D structures	PDB ID or PDB file	All RNA sizes	Table with function, PDB ID, 3D alignment and experimental method	Several seconds

RESULTS AND DISCUSSION

Algorithms, Inputs & Outputs

A. ARTS (Alignment of RNA Tertiary Structures)

ARTS which was made in 2005 by Dror et al. is a computational method to compare and identify common RNA 3D substructures [7]. It can be accessed simply through http://bioinfo3d.cs.tau.ac.il/ARTS/. The algorithm of ARTS uses RNA tertiary (3D) structure as the basis because it was thought that secondary structures won't be able to correctly predict the actual tertiary structure of the RNA.

The algorithm that they use to compare if two RNA structures are similar or not uses a combination of pattern-matching and hashing algorithm. First, they separate one of the RNA structures into different groups of two adjacent base pairs (termed "k" in this article) found in that structure. Each k is then put into a hash table which will be extracted by the other RNA structure considered as the query. Each k match between the two RNA structure is called a seed match and is further aligned by the RMSD (Root Mean Square Deviation) value.

This algorithm is also used to build their database from RNA structures available in the PDB database. They use this algorithm to compare between all available 1D, 2D, & 3D RNA structures in the PDB at that time and inserted only non-redundant results into their database. This causes them to have a considerable small sized database – compare to the other tools mentioned in this article – with only 244 RNA structures stored.

To use this search tool for an RNA structure, the user can simply input the structure in the from of a PDB file and the tool will run the algorithm to search for similar RNA structures. Once the program has finished, the results will be emailed to the user's provided email and the user can open the link to look at the output. The output format itself is fairly simple. A table that contains the PDB ID of the similar RNAs, number of nucleotides & base pairs in each RNA, the RMSD value, and a Jmol link to view the 3D format of the alignment.

B. FASTR3D

FASTR3D, which stands for "A Fast and Accurate Search Tool for RNA 3D Structures", was created by Lai et al. in 2009 as a web-based tool to search for similar RNA structures [9]. The link to access the tool itself is http://genome.cs.nthu.edu.tw/FASTR3D/. FASTR3D uses an algorithm based on RNA secondary (2D) structures to build its database and search algorithm unlike ARTS. The reason is quite the opposite of ARTS. Similar tertiary structures are thought to be difficult to compare because it is hard to find a constant ratio approximation between them. With that said, FASTR3D is commonly used as a benchmarking tool for new programs that aims to find similar RNA 3D structures.

FASTR3D uses a simpler algorithm to build its database and search engine compared to the other tools mentioned in this article. The algorithm is divided into three main steps. The first step is to build the database itself using a form of hashing algorithm. Primary sequence, secondary structure, and tertiary structure of all RNAs in the PDB database are all derived into secondary structures consisting of Watson-Crick and wobble base pairs. These derived structures are then stored in a hash table that acts as the tool's database. Once the database has been made, the second step is to analyze the user's input.

The user can input RNA 3D structure data in three ways; a RNA's tertiary structure in the form of a PDB code and specified residue range, a RNA's secondary structure in dot-bracket notation, and a FASTA format of a RNA's primary sequence. The algorithm is then rerun to derive the user's input into a secondary structure like in the database. The third and final step of the algorithm is to search through the whole database to find structures that match the input RNA. There is an option for the user to use a primary sequence or tertiary structure filter to rule out matches that isn't equal to the primary sequence and/or tertiary structure.

The output format of FASTR3D is a table consisting of matches that was found in the database. It includes data such as PDB ID, chain ID, primary sequence, secondary sequence, tertiary sequence (in the form of a Jmol link), residue range, RNA class, and experimental method. RNA class refers to the function or the type of RNA (i.e. rRNA, tRNA, etc). The experimental method in this case is the method used to verify the structure and/or function of the RNA using in vivo or in vitro techniques.

C. RNA FRABASE 2.0

RNA FRABASE 2.0 is the updated version of RNA FRABASE (RNA Fragments Search Engine & Database) which was created by Popenda et al. in 2010 [10]. It can be accessed using the link http://rnafrabase.cs.put.poznan.pl/index.php?act=examples&act2=.It is a search engine used to find RNA 3D substructures with its own database based on RNA secondary (2D) structures. It has similarity to FASTR3D in terms of the search algorithm, however, the database itself is created with a different algorithm.

RNA FRABASE builds their database in a similar way to FASTR3D by deriving RNA structure data obtained from the PDB database. Unlike FASTR3D, however, the derivation is based on the atom coordinates of all RNA structures deposited in PDB. This allows them to store more as the derivation separates data not only based on secondary structure, but also on primary sequence, torsion angles, base-base parameters, and the atom coordinates itself.

The searching algorithm for FRABASE 2.0 itself is not that different from its previous version. It still uses a pattern-searching algorithm based on the RNA fragment inputted by the user [11] that's been updated. The pattern-searching algorithm in FRABASE 2.0 has a different threshold which allows the user to use small fragments as input and more filtering options to narrow down the match results.

The only requirement needed for the user is to input either a PDB file or manually input a primary sequence or secondary structure in dot-bracket notation.

RNA FRABASE 2.0 creates an output result much like FASTR3D using a table summary which includes PDB ID, chain ID, primary sequence, secondary structure, tertiary structure (Jmol link), residue range, RNA class, and experimental method. It also includes an additional column for the NDB ID of the RNA structures.

D. RAG-3D (RNA-As-Graphs-3D)

RAG-3D is a web-based database and search tool for similar RNA 3D structures and substructures that was created in 2015 by Zharan et al. [5]. It can be accessed under http://www.biomath.nyu.edu/?q=RAG3D. Similar to ARTS, RAG-3D uses RNA tertiary structures to build its database with a more sophisticated algorithm. The main reason is quite similar to ARTS as well, where using RNA 2D structure can become less accurate in predicting the actual tertiary motifs of RNA because they usually don't consider indels or mutations and only focus on the folding itself [7].

RAG-3D uses an algorithm where the RNA 3D structures obtained from PDB are transformed into 3D tree graphs categorized based on the connectivity of the 2D graphs made in the previous version (RAG). These graphs are then further cataloged based on the number of vertices in the graph. The RAG-3D database have a total of 36 graphs and subgraphs between 2-10 vertices stored in their database. Each graph can contain several RNA 3D structures at the same time based on the similarity of those structures when turned into a 3D graph. This is then used to build their search algorithm for the online user interface of the website.

The RAG-3D search algorithm searches for graph similarity between the query graph and subgraphs to the graphs available in the database. The graph similarity is determined based on the connectivity pattern and RMSD value between the query and database 3D graphs. The user can input the query RNA in the form of a PDB ID or a PDB file with the chain ID as an optional input.

RAG-3D creates an output format for the user separated in two parts, the query information and the similarity result. The query information contains all of the secondary motifs and subgraphs extracted from the query sequence as well as the 3D representation of the top result, while the similarity result is a table that contains all the information of the similar structures and substructures ranked based on the RMSD values. It also includes the PDB ID, RMSD values, corresponding function and experimental method of the substructure according to the PDB data. The corresponding function column refers to the same column as RNA class in FASTR3D [9] and RNA FRABASE 2.0 [10].

E. R3D-BLAST2

R3D-BLAST2 is a similar tool that comes after RAG-3D and is created in 2017 by Yen et al. [6]. It seems to be the newest tool made so far and is also a database-search tool for RNA 3D substructures that is built from RNA tertiary (3D) structures data from the PDB database which can be accessed from http://140.114.85.168/R3D-BLAST2/. However, R3D-BLAST2 differentiates itself from RAG-3D by using an entirely different approach in creating its database and search algorithm.

The algorithm used in R3D-BLAST2 to build its database is a structural-alphabet (SA) based algorithm where the RNA 3D structure data they obtain from the PDB database are clustered together based on their nucleotides into 23 conformation clusters. These conformation clusters are each given a capital letter to become a 1D SA sequence and the structures are stored based on their similarity to these 1D sequences when transformed.

For the search algorithm, they utilize a modified form of BLAST (sequence alignment) to search the SA-encoded sequences stored in the database with the query structure given by a user. The ranking of the results is based on the E-value of each hit. The type of input the user can give is either a PDB ID or a PDB file and the chain ID for a more detailed search. However, since R3D-BLAST2 is an upgraded version of R3D-BLAST, the user can also further filter out the results based on RMSD values, SAS (Structural Alignment Score) values, and/or PSI (Percentage of Structural Identity) values. The three filters are not in the default options but can be edited in the advance parameter section.

The output format of R3D-BLAST2 is similar to RAG-3D in terms of the list of similar 3D substructures given in a table. While R3D-BLAST2 also has function and experimental method in the result table, they also output the E-value and query coverage of each results. There is also a JSmol link in each result to show the alignment result in 3D format.

Performance Evaluation

To compare between all the tools mentioned in this article, a performance evaluation was done for several categories. The categories tested will be based on time, accuracy, database relevance, and overall performance. To evaluate each category, each tool was tested using three different kinds of RNAs under default parameters (unless specified). A summary of the type, chain, and length of the RNA can be seen in Table 2.

Table 2 Sample Dataset for tool evaluation.

RNA Type	PDB ID	Chain ID	Residue Length
Pseudoknot	1HVU	C	30
5S rRNA	3CC2	9	122
18S rRNA	3J80	2	1799

A. Time

When tested, ARTS took the longest time to process each RNA sample with an average time of 350.1 seconds (~5.8 minutes). Both FASTR3D and FRABASE, on the other hand, was found to be the fastest tool in processing all of the samples. Each tools processing time are averaged from the three samples and can be seen in Table 4. However, considering the database and algorithm used for each tool, the time taken to process the RNA sample does not mean the tool itself is efficient or accurate.

With ARTS, the algorithm it uses where it takes each adjacent base pair in the query RNA and compare it to adjacent base pairs of RNAs in their database stored as hash table can be considered as the main reason for its slow performance. The tool uses a very greedy algorithm approach in order to produce its results where it compares everything one-by-one and adding each match to the table of result. Another reason could also be because the result is sent to the user through email instead of being given in the same tab or window. Depending on the internet speed of the user, the result could be received at a much slower rate than it was supposed to. It doesn't seem like ARTS performance based on its speed can be compared to the other four tools.

When comparing FASTR3D & FRABASE, however, there seems to be some similarity. Both tools use RNA 2D structure as their base and seem to process the sample in almost the same amount of time when being averaged, 3.6 & 2.1 seconds respectively. Due to the algorithm of both tools only considering the secondary structure of the query RNA and not the final tertiary structure, both FASTR3D & FRABASE wouldn't consider any form of mutations like insertions and deletions when processing the query RNA. This allows them to search through their database faster as they would only search specifically to the query RNA without modifying anything.

The final two, RAG-3D & R3D-BLAST2, have slightly different average time when processing the three RNA samples. RAG-3D seems to have a slightly longer average at 26.5 seconds, while R3D-BLAST2 has an average of 14 seconds.

B. Accuracy

The accuracy of each tool is determined by the final results obtained from each RNA sample. Since the samples used are known samples and not novel RNA structures, the result should reveal the sample as the top matching similar RNA structure. This determines how effective the algorithm used not only to search the structures, but also when building the database of each tool.

As seen in Table 3, each tool produced different amount of results the larger the sample size. For the pseudoknot sample, most of the tools acquired the same amount of results. Although RAG-3D's results has more than FASTR3D/FRABASE/R3D-BLAST2, the pseudoknot sample itself was found to be the highest matching result in all four tools. RAG-3D has more results mainly because the database isn't based on the original structure but on the 3D graph representation. As such, if the 3D graph of the query RNA is found to be similar to a certain group in the database, then the result would be all the RNA structures in that group.

For the 5S rRNA results, FASTR3D and FRABASE produced similar amounts, but much lower than R3D-BLAST2. This is because the database size of the tools based on 2D is smaller than R3D-BLAST2, with FASTR3D having the least amount of RNA structures than the other two. Another reason is also because using 2D structures to compare similarity means they have opted to ignore possible insertions or deletions in the query structure, making the program consider less parameters when searching for matching results. Meanwhile, RAG-3D couldn't produce any results due to the server being timed out from running the sample for too long, even though the sample seemed to be available in their 3D graph database. It could be the case that the query size is bigger than the ones stored in the database, where they've separated it into subgraphs. However, with that said, they all still produce the query sample as the highest matching result proving their accuracy with a larger query sample.

However, three of the tools (FASTR3D, FRABASE, RAG-3D) doesn't produce any result for the 18S rRNA and only R3D-BLAST2 seemed to have no problem in finding matching structures. This is due to the unavailability of the data in their databases. RAG-3D's database is limited to the number of vertices each 3D graph group has and with the maximum graph only having 10 vertices, it limits the database in processing very large samples. In this case, the 18S rRNA sample used seems to create a big 3D graph representation making it impossible for the algorithm to read it hence why it produced no results. As for FASTR3D & FRABASE, their database only stores around 1,300 & 2,700 RNA structures from PDB, respectively, and it seemed like it hasn't been updated since the creation of the databases. Seeing as the 18S rRNA sample was released after the tools' creation, it was found that both tools didn't have the RNA sample in their databases making it impossible for the algorithm to find any matches.

Out of all the tools, ARTS seems to be the only outlier, producing similar amount of results for each RNA sample. The smallest sample had 171 matches, while the other two samples had 239 & 240 matches, respectively. Since the algorithm matches the sample with the database based on adjacent base pairs, it would be possible for the algorithm to consider even different sized structures to be similar. In this case, even the smallest sample would produce a lot of matches. Even though the pseudoknot result had the sample as the highest match, the 5S & 18S rRNA samples didn't have the sample as the matching result at all. With only 244 RNA structures stored in the database, it's possible that the sample is non-existent. This makes ARTS the tool with the worst accuracy in finding similar RNA structures.

Table 3 Search Results for each RNA sample.

Tool Name	Pseudoknot	5S rRNA	18S rRNA
ARTS	171	239	240
FASTR3D	4	20	0
RNA FRABASE 2.0	4	26	0
RAG-3D	10	Timed out	0
R3D-BLAST2	4	757	8,727

C. Database Relevance

The last category is used to determine whether the database created by each tool is still considered relevant or not. This is determined by the last time database was updated and the ability of the database to process newly released samples as can be seen in Table 4. Starting from the oldest tool, the database of ARTS seems to not have been updated since its initial creation and it was mentioned in the webpage that the data was collected from PDB in March 2007. With only a final storage of 244 structures and an outdated repertoire of structures, ARTS has the biggest difficulty in analyzing newly released RNA structures.

This is also a similar case with FASTR3D & FRABASE. With only around 1,300 in FASTR3D & 2,700 in FRABASE collected from PDB, they would produce results with a much lower number than they should in the present time. Even though they have a much better accuracy than ARTS, they would only produce results based on data available in their database for novel query samples and wouldn't be able to include structures that was released after the creation of their databases.

RAG-3D's relevance is limited by the algorithm's capacity to read bigger RNA structures as it can only read graphs with up to 10 vertices. This makes the relevancy of the tool somewhat ambiguous as it can still process newly released structures that's in the size range of the algorithm's capacity, but any novel structures that has a graph of more than 10 vertices would produces an error as the program can't read it. Meanwhile, R3D-BLAST2 being a tool created in 2017 seems to be the most relevant as the database was created using PDB data from March 2017. With also having no problem in analyzing large sized samples, R3D-BLAST2's ability to process novel RNA structures would yield the most results with great accuracy and efficiency.

Table 4 Category evaluation of each tool.

Tool Name	Accuracy	Average Time Needed*	Database Relevance
ARTS	Not accurate	350.1 sec	Not updated since 2007
FASTR3D	Accurate (no indels)	3.6 sec	Not updated since 2009
RNA FRABASE 2.0	Accurate (no indels)	2.1 sec	Not updated since 2010
RAG-3D	Accurate (depending on input size)	26.5 sec	Updated based on algorithm capacity
R3D-BLAST2	Accurate	14 sec	Updated regularly

*: Calculated from the three RNA samples used

D. Overall Performance

Looking at the results for each category, the overall performance can be determined with small competition between each tool. The processing time, accuracy of the results, and relevance of the database to recent findings, determine how well the tool works as a whole. In each category, there is quite a big gap between the results from each tool that can be observed. With ARTS, it performs poorly as an overall due to being not only the oldest tool than the other four, but also using a greedy algorithm to analyze the sample making it take a longer time to process everything and having the least amount of accuracy between the query and the database.

FASTR3D & FRABASE is somewhat outdated with a database that's not updated in the last 10 years. However, the speed and accuracy of their algorithm makes them still a regularly used tool to determine the quality of newly created tools with similar purposes. As an example, both RAG-3D and R3D-BLAST2 used them to determine the quality of their tool when it was initially created. Although outdated, their performance is still considered excellent and is sufficient enough to be used as a common benchmarking standard

RAG-3D's quality as a search tool for RNA 3D structures is considered high enough and can be compared to other tools previously made or will be made without falling behind. It may have a slightly slower processing time than R3D-BLAST2, however, the accuracy of the search results is on par. It also provides the user with information of possible secondary motifs (folding, loops, etc.) from the query

RNA structure, giving the user a better understanding on how the algorithm analyzes and produces result for that specific sample. However, the size of the RNA sample should be considered since the tool can only analyze RNA structures up to a certain size.

R3D-BLAST2 is the tool with the best overall performance based on the category evaluation. With a short processing time, producing highly accurate results, and a database that is still relevant (2-year gap), R3D-BLAST2 would be the best recommended tool to find similar RNA 3D structures and/or substructures. Its benchmarking against older tools like FASTR3D & FRABASE also produces satisfactory results without it falling behind in the other category.

CONCLUSION

In the last decade, information of RNA 3D structures stored in databases such as PDB and NDB have been increasing exponentially making it a necessity to build tools for finding similar RNA 3D structures in order to help the annotation process. By benchmarking and comparing tools previously made, a better understanding for future approaches can be made. The results have shown that accuracy and efficiency of the algorithm to search for similar RNA 3D structures is one of the essential necessities in creating these tools. Not only that, the result of the search is also highly dependent on the database of the tool itself. Database that are not updated will cause the tool to have lower accuracy in the future. So far, R3D-BLAST2 has proven to be the tool with the best quality in the present. In the future, it would be wise to consider a search tool with a database that updates regularly following PDB and/or NDB.

ACKNOWLEDGEMENT

The authors of this article would like to thank the Institute of Research and Community Empowerment (LPPM), Department of Bioinformatics, and Center of Student Development Department of Indonesia International Institute for Life Sciences (i3L) for providing this opportunity. Lastly, thanks also goes to Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia for providing Hibah Penelitian Dasar DIKTI/LLDIKTI III 2019 No. 1/AKM/PNT/2019. The author declares that there is no conflicting interest.

REFERENCES

[1] Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2(12), 919–929.

[2] Mattick, J. S., Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, 15(1), R17–R29.

[3] Rose, P. W., Prlić, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., Young, J., Zardecki, C., Berman, H. M., Bourne, P. E., Burley, S. K. (2015). The RCSB Protein Data Bank: views of structureal biology for basic and applied research and education. Nucleid Acids Research, 43, D345–D356.

[4] Narayanan, B. C., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., Leontis, N. B., Berman, H. M. (2014). The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Research*, 42, D114–D122.

[5] Zahran, M., Bayrak, C. S., Elmetwaly, S., Schlick, T. (2015). RAG-3D: a search tool for RNA 3D substructures. *Nucleic Acids Research*, 43(19), 9474–9488.

[6] Yen, C. Y., Lin, J. C., Chen, K. T., Lu, C. L. (2017). R3D-BLAST2: An improved search tool for similar RNA 3D substructures. BMC *Bioinformatics*, 18(S16), 574.

[7] Dror, O., Nussinov, R., Wolfson, H. (2005). ARTS: alignment of RNA tertiary structures. *Bioinformatics*, 21(S2), ii47–ii53.

[8] Abraham, M., Dror, O., Nussinov, R., Wolfson, H. J. (2008). Analysis and classification of RNA tertiary structures. RNA, 14(11), 2274–2289.

[9] Lai, C. E., Tsai, M. Y., Liu, Y. C., Wang, C. W., Chen, K. T., Lu, C. L. (2009). FASTR3D: a fast and accurate search tool for similar RNA 3D structures. *Nucleic Acids Research*, 37, W287–W295.

[10] Popenda, M., Szachniuk, M., Blazewicz, M., Wasik, S., Burke, E. K., Blazewicz, J., Adamiak, R. W. (2010). RNA FRABASE 2.0: An advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics*, 11(1), 231.

[11] Popenda, M., Blazewicz, M., Szachniuk, M., Adamiak, R. W. (2007). RNA FRABASE version 1.0: An engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Research*, 36, D386–D391.