Published online 29 May 2011 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* ***W45–W49***
doi:10.1093/nar/gkr379

R3D-BLAST: a search tool for similar RNA 3D substructures

Yun-Chen Liu[1,2], Chung-Han Yang[1,2], Kun-Tze Chen[4], Jyun-Rong Wang[1,2], Mei-Ling Cheng[1,2], Jen-Chun Chung[1,2], Hsien-Tai Chiu[2,3,*] and Chin Lung Lu[4,*]

[1]Institute of Bioinformatics and Systems Biology, [2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 300, Taiwan, [3]Department of Chemistry, National Cheng Kung University, Tainan City 701 and [4]Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan

Received March 3, 2011; Revised April 25, 2011; Accepted May 1, 2011

ABSTRACT

R3D-BLAST is a BLAST-like search tool that allows the user to quickly and accurately search against the PDB for RNA structures sharing similar substructures with a specified query RNA structure. The basic idea behind R3D-BLAST is that all the RNA 3D structures deposited in the PDB are first encoded as 1D structural sequences using a structural alphabet of 23 distinct nucleotide conformations, and BLAST is then applied to these 1D structural sequences to search for those RNA substructures whose 1D structural sequences are similar to that of the query RNA substructure. R3D-BLAST takes as input an RNA 3D structure in the PDB format and outputs all substructures of the hits similar to that of the query with a graphical display to show their structural superposition. In addition, each RNA substructure hit found by R3D-BLAST has an associated *E*-value to measure its statistical significance. R3D-BLAST is now available online at http://genome.cs.nthu.edu.tw/R3D-BLAST/ for public access.

INTRODUCTION

Like proteins, finding structural similarities between RNAs often helps biologists to better understand their shared potential functionality that might not be detected only using RNA primary sequences. As RNA structures currently solved and deposited in the Protein Data Bank (PDB) (1) continue to grow in both number and size, exhaustive search approaches to comparing a query RNA structure with every RNA structure in the PDB can be very difficult and time consuming because finding their similar substructures has been shown to be computational intractable (2). Therefore, currently available search tools, such as RNA FRABASE (3,4) and FASTR3D (5), employ some heuristics to search the PDB for fragments of known RNA 3D structures that possess the same 2D structures as found in the entire query RNA, which can be implemented in an efficient way. However, these search tools still suffer from some limitations as described below, although they can serve as useful tools to process and analyze various aspects of RNA 3D structures in the PDB.

Basically, both RNA FRABASE and FASTR3D, as mentioned above, are mainly dedicated to searching for RNAs whose 2D structures annotated in the PDB are entirely identical to that of the query RNA. In particular, they do not allow insertions and deletions in their structural alignments for comparing the query with each RNA structure in the PDB. However, there are many RNAs whose overall 3D structures are similar, but their entire 2D structures, as well as their primary sequences and lengths, are not the same. For instance, the two tRNAs (PDB IDs: 1EHZ and 3KIX) shown in Figure 1 exhibit similar whole 3D structures, but they have different sequences, lengths and 2D structures annotated in the PDB. Therefore, while querying one of these two tRNAs, RNA FRABASE and FASTR3D both fail to find the other one with any structural similarity. On the other hand, some RNAs may share only similar 3D substructures, rather than entire 3D structures, as illustrated in Figure 2. Due to the intrinsic limitations described above, both RNA FRABASE and FASTR3D will miss these locally similar RNA structures in their search results. In addition, neither of RNA FRABASE and FASTR3D currently provides any statistical values (e.g. *E*-value) for each RNA structure hit to indicate its significance.

In this study, we have therefore developed a BLAST-like search tool, called R3D-BLAST, to overcome the

*To whom correspondence should be addressed. Tel: +886 3 5731205; Fax: +886 3 5723694; Email: cllu@cs.nthu.edu.tw
Correspondence may also be addressed to Hsien-Tai Chiu. Tel: +886 3 5131595; Fax: +886 3 5719605; Email: chiu@mail.nctu.edu.tw

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

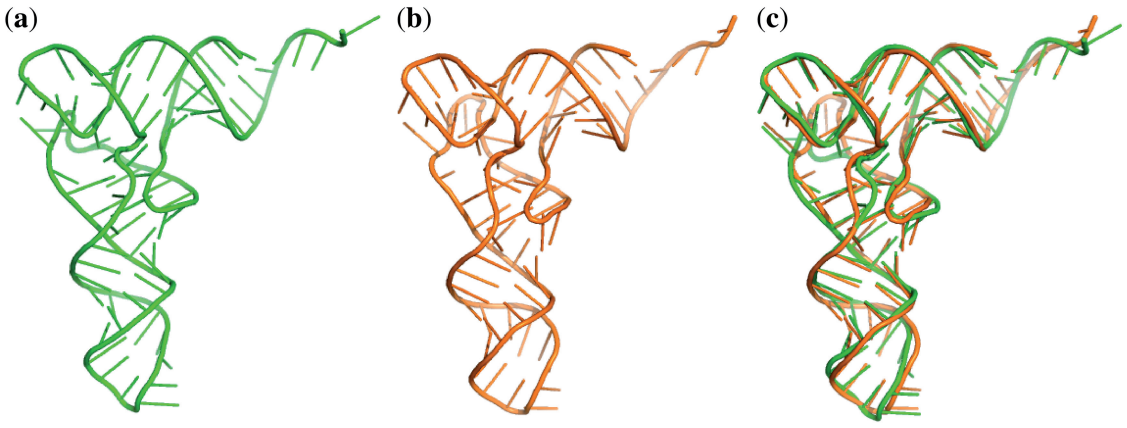

Figure 1. Two tRNAs: (**a**) PDB ID: 1EHZ and chain ID: A; (**b**) PDB ID: 3KIX and chain ID: W; (**c**) superimposition of their 3D structures with Root Mean Square Deviation (RMSD) of 2.457 Å; and (**d**) their 1D sequences (with identity of 44%) and 2D structures annotated in the PDB.

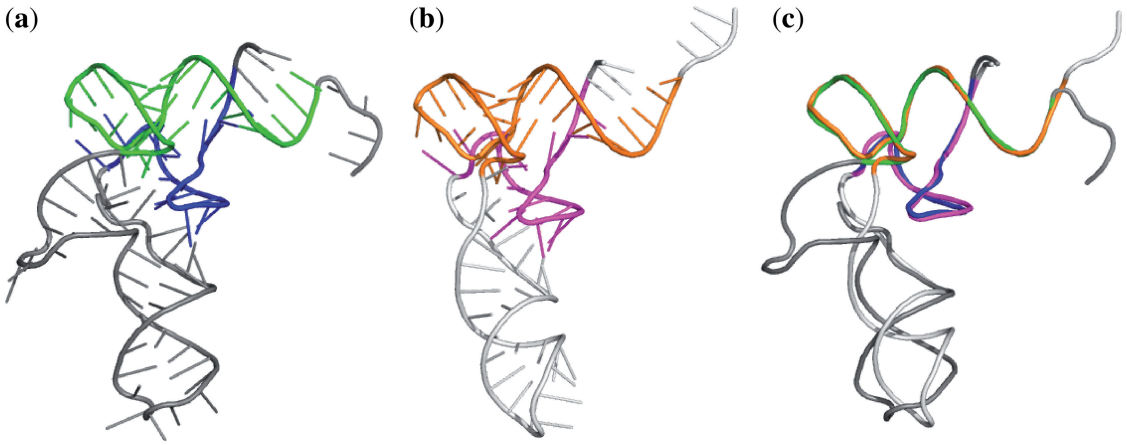

Figure 2. Two tRNAs: (**a**) PDB ID: 1WZ2 and chain ID: D; (**b**) PDB ID: 3KIX and chain ID: V; and (**c**) their two similar substructures, one in blue and magenta (with an RMSD of 1.314 Å) and the other in green and orange (with an RMSD of 1.622 Å).

above limitations in RNA FRABASE and FASTR3D. The basic idea behind R3D-BLAST is that all the RNA 3D structures in the PDB are first encoded as 1D structural sequences using a structural alphabet developed in our previous work (6), and BLAST, a popular bioinformatics tool to find homologous proteins/RNAs only based on their sequence similarity (7), is then applied to search these 1D structural sequences for the RNAs whose 1D structural sequences are locally similar to that of a query RNA. The experimental results have shown that our R3D-BLAST (i) can quickly return its search result to the user, (ii) has better performance than both RNA FRABASE and FASTR3D for finding those RNAs whose entire or nearly entire 3D structures are similar to that of the query RNA and (iii) outputs a lot of other RNAs that have substructures similar to that of the query, which, however, cannot be done either by RNA FRABASE or FASTR3D. In addition, our R3D-BLAST provides an *E*-value for each RNA substructure hit to indicate its statistical significance.

It is noteworthy that, although our R3D-BLAST is dedicated to searching for similar RNA substructures, it can still be used to find some RNA structural motifs that particularly are single-stranded, such as hairpin loops and pseudoknots, where their nucleotides are contiguous in the RNA sequence. Actually, several useful tools, like PRIMOS (8) and FR3D (9), have currently been developed to search for a wide variety of motifs in RNA structures, including multi-stranded ones like *K*-turn, *S*-turn and sarcin-ricin motifs. PRIMOS uses two pseudo-torsion angles to represent RNA backbone conformations and performs structural motif searches by analyzing the differences in the backbone conformation between a query motif and the RNA structures

in the PDB. As for FR3D, it searches for a query motif in an RNA structure based on a base-centered approach that analyzes geometric discrepancy between the query and candidate motifs and/or uses symbolic constraints specifying base–base interaction, base identity and sequence continuity.

METHODS

The algorithm we used to implement our R3D-BLAST is as follows. We first encode all the RNA 3D structures currently deposited in the PDB (as of December 2010) as 1D structural sequences using the structural alphabet of 23 distinct nucleotide conformations, which was constructed previously by us using the pseudo-torsion angles (η and θ) of RNA nucleotide backbones (6). We then apply BLAST (7) to searching these 1D structural sequences for the RNA substructures whose 1D structural sequences are similar to that of the query RNA substructure. To properly score each alignment obtained by R3D-BLAST, we also define a 23×23 BLOSUM-like substitution matrix based on the observed and estimated probabilities of occurrence for each pair of aligned structural letters (6). In addition, our R3D-BLAST provides an *E*-value to measure the statistical significance of each RNA substructure hit. The *E*-value is obtained by calculation with the Karlin–Altschul equation $E = Kmn^{-\lambda S}$ (10), where m and n, respectively, are the sizes of the query RNA structure and the database, S is a raw score of the structural alignment and K and λ are statistical parameters, respectively, depending on the structural letter composition of RNA structures being aligned and the used scoring function. To obtain accurate parameter estimates, we utilize the island method, which was proposed by Altschul *et al.* (11) for estimating our K and λ values, according to different sets of gap open and extension penalties. It is inevitable that some RNA substructure hits returned by R3D-BLAST, particularly with high *E*-values, may not be similar to the substructure of the query RNA in their 3D structures. Therefore, we design an optional filter for further screening out the RNA substructure hits whose RMSD values and/or structural alignment scores (SAS) with respect to the query RNA substructure are greater than some predefined thresholds. SAS is defined as $100 \times \text{RMSD}/(\text{number of aligned residues})$, which was introduced by Kolodny *et al.* (12), who have shown that SAS can serve as a useful measure to separate good structural alignments from less good ones. In the interest of running time, the RMSD/SAS filter in R3D-BLAST is not selected by default.

USAGE OF R3D-BLAST

Except the BLAST program that was downloaded from NCBI, the kernel of R3D-BLAST, as well as its web interface, was written in PHP. The server of R3D-BLAST is currently installed on IBM PC with 2.8 GHz processor and 3 GB RAM under Linux system and its RNA structure database is updated every 3 months. R3D-BLAST is easy to operate for the user either by entering a PDB ID for the query RNA structure or by uploading its file in the PDB format, optionally with specified residue range. Then R3D-BLAST will quickly return its search results to the user in an output page that contains the information of each RNA hit in the PDB, including its PDB ID, chain ID, title to describe the experiment, classification of molecule, experimental method to determine the structure, structure resolution and released date, and the details of its structural alignment with the query, including *E*-values, range and length of aligned residues and calculated RMSD and SAS (if RMSD/SAS filter is selected). In the output page, R3D-BLAST also provides the user with a hyperlink to show the graphical display of the structural superposition between the query and each R3D-BLAST hit by Jmol, as well as the details of their structural sequence alignment and primary sequence alignment. If needed, the user can select different parameter settings about alignment (that can be either gapped or ungapped) and its used gap penalty to run R3D-BLAST, and also modify the *E*-value threshold to further restrict the search results of R3D-BLAST. The snapshot of running our R3D-BLAST is shown in Figure 3.

EXPERIMENTAL RESULTS

To validate R3D-BLAST, we have tested it on some RNA 3D structures, and also compared its search results with those obtained by similar tools, RNA FRABASE (version 2.0) and FASTR3D, even though RNA FRABASE and FASTR3D are both designed for searching similar whole RNA structures while our R3D-BLAST is for similar RNA substructures. Unless specified, all the tools were run on the respective servers with default parameters. The experimental results we obtained by testing R3D-BLAST, RNA FRABASE and FASTR3D on five different kinds of RNAs with different lengths are shown in Table 1, where the percentage in the title of the table means the query coverage, which is defined as the percent of the query length that is included in the structural alignment. Except tRNA in this testing dataset, our R3D-BLAST found the same or more number of RNA hits with 3D structures entirely similar to that of the query (i.e. with query coverage of 100%) as compared to both RNA FRABASE and FASTR3D. This is because that, as mentioned before, two RNAs may share a very similar whole 3D structure even though their 2D structures, as well as their sequence lengths, are different. As for the tRNA, both RNA FRABASE and FASTR3D returned more RNAs when compared with the RNA hits with 100% query coverage in the search result of our R3D-BLAST. Recall that the algorithms in both RNA FRABASE and FASTR3D were designed to search for those RNAs whose entire 2D structures are exactly the same as that of the query. Indeed, those RNAs found by RNA FRABASE or FASTR3D have the same 2D structures. However, some regions near the ends of these tRNA 3D structures are not quite similar. As a result, the dissimilar regions were finally removed from the tRNAs returned by our R3D-BLAST. In fact, all the tRNAs identified by both RNA FRABASE

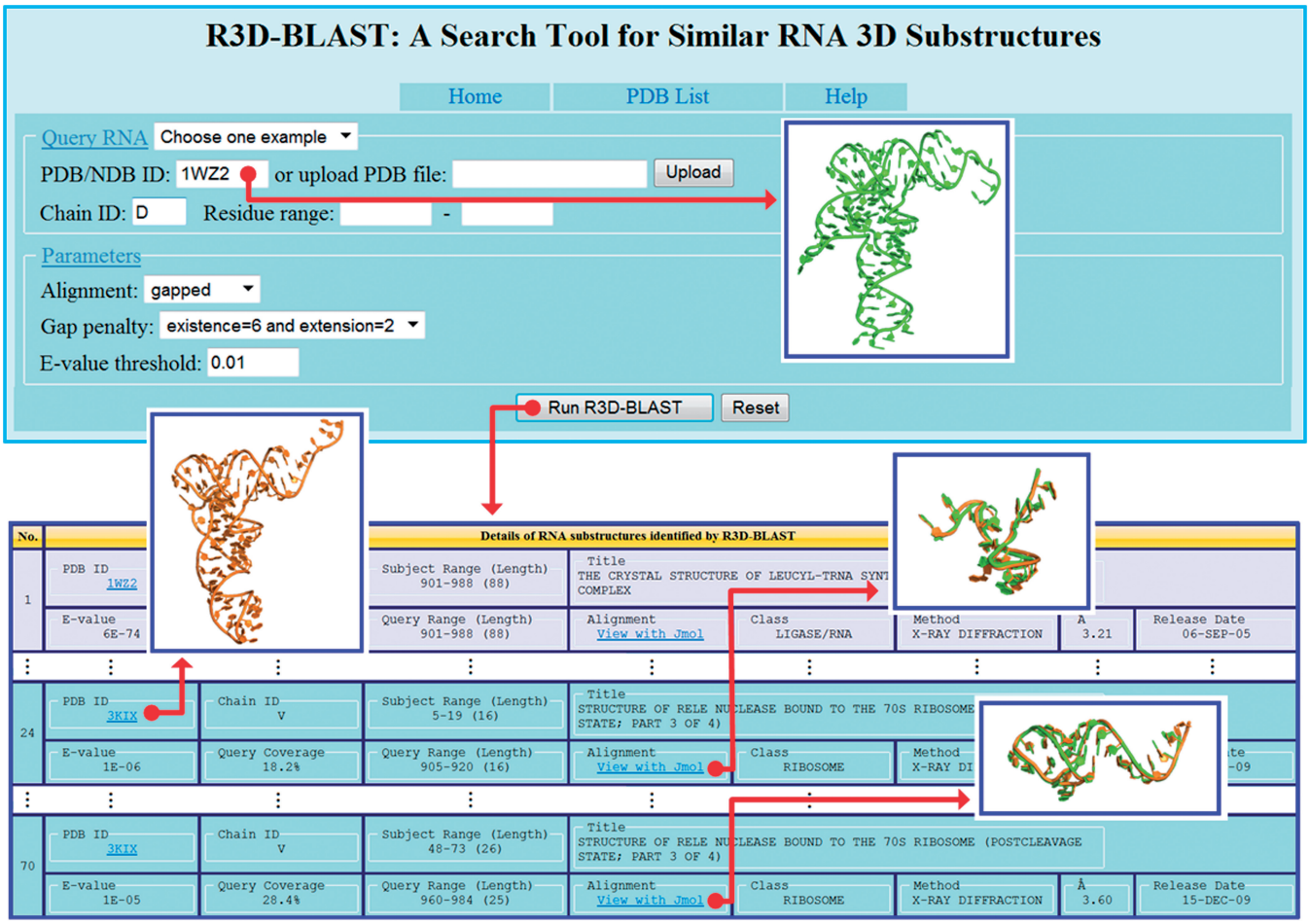

Figure 3. The top panel shows the interface of R3D-BLAST with a query of tRNA (PDB ID: 1WZ2 and chain ID: D), and the bottom panel presents its search results by highlighting two similar substructures identified in another tRNA (PDB ID: 3KIX and Chain ID: V) and their superpositions with the corresponding substructures of the query.

Table 1. Comparison of search results obtained by R3D-BLAST, RNA FRABASE (version 2.0) and FASTR3D

Query	PDB_Chain	Range/length	R3D-BLAST				RNA FRABASE	FASTR3D
			100%	≥90%	<90%	Total		
Pseudoknot	2AVY_A	499-544/46	89	138	11	149 (2.4)	83 (0.7)	40 (0.6)
Riboswitch	1Y27_X	14-81/68	3	9	18	27 (1.2)	1 (0)	1 (0)
tRNA	1EHZ_A	1-76/76	7	62	259	321 (3.1)	15 (1.6)	17 (1.3)
5S rRNA	3CC2_9	1-121/121	40	63	184	247 (2.3)	19 (0.2)	17 (0.2)
Ribozyme	1X8W_B	96-414/247	1	1	15	16 (2.1)	1 (0)	1 (0)
16S rRNA	2AVY_A	5-1534/1530	2	8	1199	1207 (4.0)	2 (0.1)	0 (0)

The values presented in the parentheses for each tool are the average RMSD values (Å) between the query and hits.

and FASTR3D, along with other tRNAs, can still be found by our R3D-BLAST with the query coverage of at least 90%. As also shown in Table 1, our R3D-BLAST returned a lot of other RNAs with their substructures similar to that of the query tRNA (with <90% query coverage), which cannot be done either by RNA FRABASE or by FASTR3D. Finally, in Table 2 we show the CPU time of our R3D-BLAST when querying with the RNAs listed in Table 1. As indicated in Table 2, our R3D-BLAST can finish its jobs within a few to several seconds if the option to filter by RMSD/SAS is not selected (default); otherwise, additional time is required for computing the RMSD and SAS values between the query and all the hits returned by R3D-BLAST, which

Table 2. CPU time for running our R3D-BLAST

Query	Filtered by RMSD/SAS	
	No	Yes
Pseudoknot	6 s	1.2 min
Riboswitch	1 s	4 s
tRNA	5 s	44 s
5S rRNA	10 s	1 min
Ribozyme	3 s	7 s
16S rRNA	36 s	21.2 min

can be from a few seconds to several minutes, depending on the number of the R3D-BLAST hits and their query coverages.

SUMMARY

To the best of our knowledge, our R3D-BLAST is the first web server that can quickly and accurately search the PDB for RNAs that share similar 3D substructures with a query RNA. We believe that it can be helpful for biologists to reveal functional and evolutionary relationships of RNAs even without detectable sequence similarity.

FUNDING

National Science Council of Republic of China under grants (NSC97-2221-E-009-081-MY3 to C.L.L. and NSC99-2113-M-009-005-MY3 to H.-T. C.) in part. Funding for open access charge: National Science Council of Republic of China.

Conflict of interest statement. None declared.

REFERENCES

1. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
2. Kolodny,R. and Linial,N. (2004) Approximate protein structural alignment in polynomial time. *Proc. Natl Acad. Sci. USA*, **101**, 12201–12206.
3. Popenda,M., Blazewicz,M., Szachniuk,M. and Adamiak,R.W. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res.*, **36**, D386–D391.
4. Popenda,M., Szachniuk,M., Blazewicz,M., Wasik,S., Burke,E.K., Blazewicz,J. and 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**, 231.
5. Lai,C.E., Tsai,M.Y., Liu,Y.C., Wang,C.W., Chen,K.T. and Lu,C.L. (2009) FASTR3D: a fast and accurate search tool for similar RNA 3D structures. *Nucleic Acids Res.*, **37**, W287–W295.
6. Wang,C.W., Chen,K.T. and Lu,C.L. (2010) iPARTS: an improved tool of pairwise alignment of RNA tertiary structures. *Nucleic Acids Res.*, **38**, W340–W347.
7. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
8. Duarte,C.M., Wadley,L.M. and Pyle,A.M. (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.*, **31**, 4755–4761.
9. Sarver,M., Zirbel,C.L., Stombaugh,J., Mokdad,A. and Leontis,N.B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.
10. Karlin,S. and Altschul,S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl Acad. Sci. USA*, **87**, 2264–2268.
11. Altschul,S.F., Bundschuh,R., Olsen,R. and Hwa,T. (2001) The estimation of statistical parameters for local alignment score distributions. *Nucleic Acids Res.*, **29**, 351–361.
12. Kolodny,R., Koehl,P. and Levitt,M. (2005) Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J. Mol. Biol.*, **346**, 1173–1188.