NASSAM: a server to search for and annotate tertiary interactions and motifs in three-dimensional structures of complex RNA molecules

Hazrina Y. Hamdani[1], Sri D. Appasamy[1], Peter Willett[2], Peter J. Artymiuk[3,*] and Mohd Firdaus-Raih[1,*]

[1]School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia, [2]Information School and [3]Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

ABSTRACT

Similarities in the 3D patterns of RNA base interactions or arrangements can provide insights into their functions and roles in stabilization of the RNA 3D structure. Nucleic Acids Search for Substructures and Motifs (NASSAM) is a graph theoretical program that can search for 3D patterns of base arrangements by representing the bases as pseudo-atoms. The geometric relationship of the pseudo-atoms to each other as a pattern can be represented as a labeled graph where the pseudo-atoms are the graph's nodes while the edges are the inter-pseudo-atomic distances. The input files for NASSAM are PDB formatted 3D coordinates. This web server can be used to identify matches of base arrangement patterns in a query structure to annotated patterns that have been reported in the literature or that have possible functional and structural stabilization implications. The NASSAM program is freely accessible without any login requirement at http://mfrlab.org/grafss/nassam/.

INTRODUCTION

In RNA, just as in proteins, the 3D structure determines the functionality of the molecule. In recent years, the Protein Data Bank (PDB; http://www.rcsb.org/pdb/) (1) has seen a significant increase in the availability of RNA crystallographic structures. There were only 2 structures containing only RNA chains in 1981, 9 structures in 1991, 242 structures in 2001 and 908 structures in December 2011. But in addition to this, there are RNA–protein complex structures; thus at the end of 2011 the total number of crystallographic structures that contained RNA chains in the PDB with a resolution higher than 3.0Å stood at 1027. This number includes many complex RNA structures such as the ribosomal subunits and various ribozymes. In addition, there are numerous important nuclear magnetic resonance (NMR) structures and significant lower resolution crystal structures.

Although the number of RNA structures is lower than that of proteins, there is an urgent need for programs that are capable of analyzing them. The analysis and annotation of the tertiary motifs in the 3D PDB structures of these molecules can be difficult when limited to only the use of molecular visualization methods. One previous initiative in this area, the Non-Canonical Interactions in RNA structures (NCIR) (2), has shown that although manual curation of non-canonical interactions in RNA structures is feasible, the current volume of structures available, the manual labor-intensive processing required, and the dependence on literature may not make such an approach the most sustainable means of annotating and analyzing RNA 3D structures.

RNA structural comparison is further complicated by the difficulty in comparing RNA structures by eye to recognize local similarities on the one hand or significant differences on the other. The superficial visual similarity between adenine and guanine, and between cytosine and uracil, results in some difficulty when visually scrutinizing RNA structures for specific patterns or motifs. In such a situation, patterns of interactions may not be immediately obvious, although they may provide important clues in understanding the mechanistic details of their functions at atomic level. This problem is further compounded by the fact that minor changes in relative base orientation may result in distinct non-canonical base pairings (3). This is especially true for large, highly complex RNA structures such as the ribosomal subunits.

Methods that can compare the folding of the RNA backbone have existed for the past decade (4). Other,

more recently available programs directed at comparing RNA folds, or for searching similar RNA 3D substructures, include WebFR3D (5), ARTS (6), RNA FRABASE (7), FASTR3D (8), CLICK (9) and R3D-BLAST (10). However, these programs do not focus on, or have limited capability to investigate, the interactions and arrangements of the RNA bases. To our knowledge, the fold comparison or substructure searching programs currently available are not able to detect cases where there may be functionally significant changes in base interactions or arrangements, despite the folding of the RNA sugar phosphate backbone being either unchanged or almost unchanged. Unlike the Nucleic Acids Search for Substructures and Motifs (NASSAM) web server presented here, many of the currently available programs are only able to analyze a limited set of structures already within the server and are not able to allow users to upload their own coordinates, while others are unable to process large input files such as the ribosomal subunits, which are also within NASSAM's capabilities.

To meet this need, we developed the graph theoretical program NASSAM to search for patterns of base arrangements in RNA 3D structures (11,12). The use of graph theoretical approaches to search for and identify 3D motifs in protein structures (13) has proven to be a viable approach to the identification of clusters of amino acid residues that perform either the same or similar chemical functions despite no detectable sequence or fold similarity. In this extension to complex nucleic acid structures, the base arrangements of interest range from simple arrangements such as canonical and non-canonical base pairs, to base platforms or more complex arrangements such as triples and larger patterns. In this article, we describe the web interface that permits the use of NASSAM by biologists who wish to analyze and annotate RNA structures.

METHODS

The NASSAM algorithm

The original NASSAM standalone program, first reported in the work of Harrison *et al.* (11) and subsequently used in the annotation of base triples by Firdaus-Raih *et al.* (12), used a base interaction pattern as a query to search against a database of RNA structures. This meant that the operator had to first either create a query pattern or extract a pattern of interest from an existing structure and then convert the pattern into pseudo-atom vector matrices before executing the NASSAM search. Such a method of operation would be of limited use via a web interface despite its proven value as an exploratory investigative tool.

To deploy the NASSAM program as a web service, the original algorithm has been re-engineered to operate in reverse. Thus, in the new version of the methodology presented here, a user is able to submit a complete complex RNA structure to be searched against a database of patterns, rather than looking for a single pattern in a database of structures as previously (11,12). Patterns in this database are either sourced from literature or are theoretical patterns that have been 'hand-crafted' for accurate and optimal retrieval as described in the NASSAM Pattern Database section below. The web server processes the coordinates of the submitted input structure into a matrix containing information on the spatial relationship of pseudo-atom vectors representing the orientation of each base with respect to each other. The matrix representing the query structure is then compared against the database of pattern matrices that represent arrangements and motifs of RNA bases. This is done using the Ullmann subgraph isomorphism algorithm (14). As a totally rigid match of the base arrangement of the query to the database patterns is not realistic, the program incorporates a distance tolerance parameter that permits a controlled amount of deviation from the distance set in the query patterns. Validation tests of NASSAM performed by Harrison *et al.* (11) at the distance tolerance values of 20, 30 and 40% revealed that the program achieved the best balance of precision (91%) and recall (93%) at the distance tolerance value of 30% and this is pre-set as the default value for all NASSAM searches. The distance tolerance parameter is a key factor in optimizing or widening the search for a specific pattern. For example, we have shown that the use of high tolerances enabled the discovery of previously unreported base triple arrangements (12). Users of the web server also have the option of extending the coverage for a search by increasing the distance tolerance values for each query.

Searching for base interaction patterns with NASSAM

PDB-formatted RNA 3D structure files provide the input for NASSAM searches. There is no restriction with regard to the experimental method used to generate the PDB file. However, for NMR sourced data, there may be multiple overlapping models of the structure, and it is therefore necessary for each model to be considered individually. A screening mechanism built into the web interface prompts the user to select any one single NMR model for analysis per submission.

NASSAM has been tested with all PDB files (up to June 2011) containing RNA chains. However, because the aim of the searches is to identify motifs and interactions associated with base arrangements, the usual type of query will in most cases be an NMR structure or a crystallographic structure containing RNA chains that has been solved to a reasonably high resolution of 3 Å or higher as this will enable the atomic details of interactions to be visualized with a higher degree of confidence. The query structure is searched against a database containing known 3D base arrangement patterns that include canonical and non-canonical base pairs, patterns of triplets and other motifs as described in more detail later.

The NASSAM output lists the types of tertiary interaction or base arrangement present in the query structure (Figure 1A). Each NASSAM run will, by default, also include calculations that identify the hydrogen bonding interactions between bases in the query RNA structures (Figure 1B). As many patterns and motifs in RNA structures also include hydrogen bonding interactions in the
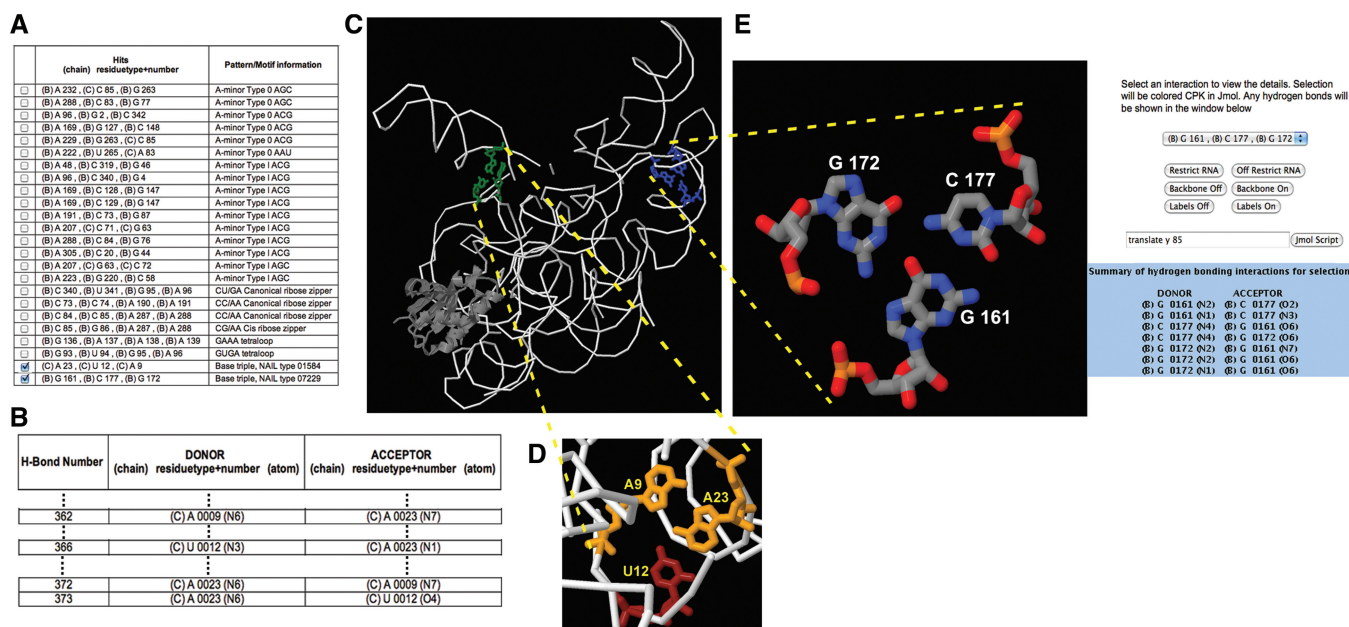
Figure 1. Snapshots of NASSAM output for the PDB structure of a bacterial ribonuclease P (PDB ID: 3Q1R): (**A**) A listing of the occurrences (hits) of base arrangement patterns or motifs; (**B**) excerpts from a list identifying hydrogen bonds between bases which in this extract is for an AUA triple; (**C**) Jmol window showing the triple hits selected in (A) relative to the whole structure; (**D**) Jmol window showing a magnification of the AUA triple selected in (A) where the hydrogen bonds formed by this interaction are presented in (B); (**E**) Jmol window showing the GCG triple selected in (A) with the hydrogen bonds present in the selection presented in an auxiliary Jmol window (right).

definition of the pattern, the generated list of hydrogen bond interactions can be used to identify whether the retrieved hits conform not only to the 3D arrangement of the bases but also to any additional parameters such as requirements for the presence of specific hydrogen bonds. In the case of the base triples, a NASSAM run is then filtered for cases where the base interactions will involve the formation of at least two hydrogen bonds per base. This filtering process will screen out numerous 'loose' opportunistic interactions involving triple-base arrangements (Tables 1 and 2) that are most likely functionally insignificant or not highly conserved. Users have a further option of selecting hit patterns for visual examination in a Jmol molecular visualization Java plug-in window (http://www.jmol.org/). Once the hits of interest have been selected, a Jmol window presenting the selected arrangements is opened (Figure 1C). Users can then browse through each base arrangement pattern using standard Jmol controls (Figure 1D) and select a specific interaction for further analysis. Any hydrogen bonding interactions that are present within the selected pattern is displayed in an auxiliary window (Figure 1E right panel).

NASSAM pattern database

The current NASSAM pattern database that is available for web searching includes patterns for both canonical and non-canonical base pairs (3,15), base triples (3,12,16), the A-minor motifs (17), kink-turn motifs (18), T-loops (19), ribose zippers (20) and tetraloops (21–23). The pseudo-atom vector representations for these interactions and motifs were mostly designed from the literature that

first presented their discoveries (as referenced previously) or via theoretical approaches (12,24). Many of the matrices representing the base pattern arrangements were 'hand-crafted' to achieve optimal retrieval and accuracy. Multiple vector combinations were tested and the representative vectors for a particular pattern was selected if it achieved optimal recall and precision when tested by searching against a dataset of PDB structures using the standalone NASSAM program. The NASSAM pattern database currently holds a total of 1041 pseudo-atom vector representations of base interaction or arrangement patterns, of which the majority are currently base triple representations. The database will be updated to include other base motifs, or newly reported base motifs.

SEARCH EXAMPLES AND CASE STUDIES

We have recently used NASSAM to perform the annotation of base triples in RNA crystallographic structures in the PDB (12). This work revealed several yet unreported triple arrangements including a novel UAU stacked triples motif that was found to be highly conserved in available structures and sequences of the prokaryotic 23S ribosomal subunits (12). In this article, we discuss specific examples regarding the utility of NASSAM versus comparable programs.

We had initially compared several programs namely: ARTS (6), WebFR3D (5), RNA FRABASE (7), FASTR3D (8) and R3D-BLAST (10) as a means of assessing the need for a web service such as NASSAM. Although these programs are intended, like NASSAM,

Table 1. Comparison of NASSAM and MC-annotate annotations for base triples in the structure 3Q1R

Base1/chain	Base2/chain	Base3/chain	NASSAM	MC-Annotate	Remarks
G4/B	A96/B	C340/B	No	Yes	Filtered out by NASSAM
G46/B	G48/B	C319/B	No	Yes	Filtered out by NASSAM
G56/B	A276/B	C325/B	No	Yes	Filtered out by NASSAM
G76/B	C84/B	A288/B	No	Yes	Filtered out by NASSAM
C73/A	G87/B	A191/B	No	Yes	Filtered out by NASSAM
A116/B	G123/B	G153/B	No	Yes	Filtered out by NASSAM
G161/B	C177/B	G172/B	Yes	Yes	–
G107/B	C209/B	A206/B	No	Yes	Filtered out by NASSAM
C231/B	G263/B	C85/B	No	Yes	Filtered out by NASSAM
A23/C	U12/C	A9/C	Yes	Yes	–
A112/B	C66/C	G19/C	No	Yes	A non-planar interaction filtered out by NASSAM
A158/B	G115/B	U154/B	Yes	No	–
G22/C	C13/C	G56/C	Yes	No	–

Table 2. Comparison of NASSAM and MC-annotate annotations for base triples in the structure 3G78

Base1/chain	Base2/chain	Base3/chain	NASSAM	MC-Annotate	Remarks
U24/A	A121/A	A247/A	No	Yes	Filtered out by NASSAM
A49/A	U59/A	G169/A	No	Yes	Filtered out by NASSAM
A67/A	A120/A	G248/A	No	Yes	Filtered out by NASSAM
G90/A	A93/A	G281/A	No	Yes	Filtered out by NASSAM
G128/A	A174/A	U238/A	No	Yes	Filtered out by NASSAM
C157/A	G220/A	G179/A	Yes	Yes	–
C127/A	G239/A	A193/A	Yes	Yes	A-minor interaction
G5/A	A260/A	G378/A	No	Yes	Filtered out by NASSAM
C6/A	G266/A	A319/A	Yes	Yes	A-minor interaction
G267/A	C286/A	G320/A	Yes	Yes	–
G320/A	U298/A	G267/A	Yes	No	Additional base-triple found by NASSAM. Overlaps with the triple G267.C286.G320 to form a base quadruple
C289/A	C358/A	G385/A	No	Yes	Filtered out by NASSAM
G288/A	G359/A	U384/A	Yes	Yes	–
A106/A	G374/A	C367/A	Yes	Yes	–
A143/A	G383/A	C360/A	No	Yes	Filtered out by NASSAM
A260/A	G378/A	U365/A	Yes	No	–
G153/A	A222/A	A154/A	Yes	No	–
A67/A	C23/A	G248/A	Yes	No	A-minor interaction
A93/A	C272/A	G281/A	Yes	No	A-minor interaction
A318/A	C7/A	G265/A	Yes	No	A-minor interaction
A60/A	G169/A	C197/A	Yes	No	A-minor interaction
A134/A	G383/A	C360/A	Yes	No	A-minor interaction

to identify RNA 3D patterns, they differ significantly from NASSAM in terms of methodology, and/or input and output. Thus, the program RNA FRABASE (7) requires input in the form of sequences or secondary structures. The WebFR3D service (5) allows for either a symbolic search of up to 15 nucleotides where the user can specify an interaction matrix for the nucleotides, or a geometric search for structures that are already available in the database, but it does not appear to allow users to upload their own structures. Others such as ARTS (6) and R3D-BLAST (10) can basically identify motifs through structural similarities achieved via either 3D or 2D alignment-based methods. The intended operation of the CLICK web service (9) is to compare and align two 3D structures and not as a direct search engine to match and identify 3D motifs. NASSAM is therefore able to fill a

void in analytical capability for RNA 3D structures that is currently not served by these programs.

The program MC-annotate (25) is, to our knowledge, the closest program to NASSAM in that it is able to perform an RNA structure annotation using a user-uploaded PDB file as input. NASSAM and MC-annotate perform the annotations using very different approaches and as a result provide generally dissimilar structural analyses. However, as described later, these can in many respects be regarded as complementary to one another. The method used by MC-annotate has been reported by Gendron *et al.* (25). NASSAM provides three output types depending on the search options selected: (i) a specific search output based on the type of interaction or base arrangement the user is interested in; or options of having all available patterns
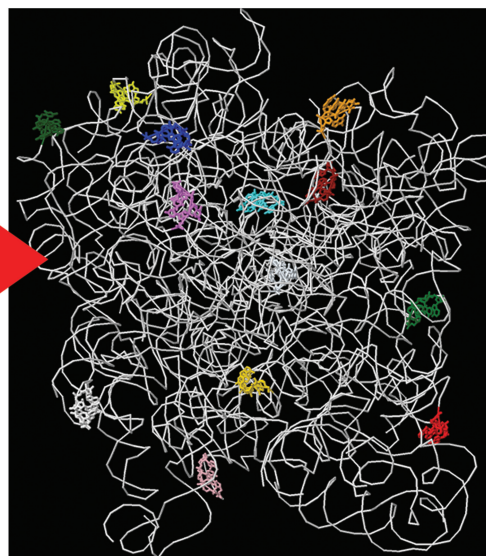
annotated either (ii) inclusive or (iii) exclusive of base pairs.

Another advantage of using NASSAM is that it is able to annotate structures according to the nomenclature used in literature and has an up to date database for these tertiary motifs. As a result, NASSAM is able to annotate motifs such as the A-minor motifs, tetraloops, kink turns and ribose zippers. A NASSAM search using the *Haloarcula marismortui* 23S rRNA subunit (PDB ID: 1FFK) as input can quickly provide a snapshot showing that there are at least 13 occurrences of the various tetraloops in the structure (Figure 2A). To demonstrate the utility of NASSAM, we compared the results of NASSAM searches against an MC-annotate annotation and here highlight the main differences in the outputs provided. We were unable to perform direct comparisons between MC-annotate and NASSAM on larger structures such as the rRNA subunits (e.g. 1FFK and 3U5D) as the MC-annotate server does not at present accept such large files for analysis, so in the examples below we focus on smaller RNA structures.

The standard NASSAM search for base triples and MC-annotate annotations for the structures of a tRNA (PDB ID: 6TNA) and a bacterial ribonuclease P holoenzyme in complex with tRNA (PDB ID: 3Q1R) were able to accurately identify the A23.U12.A9 (Figure 1) base triples in 6TNA as well as the A23.U12.A9 in the tRNA component and a G161.C177.G172 triple in 3Q1R (Table 1). NASSAM base triple searches were designed to filter out triples that were not in a planar arrangement and those that did not satisfy criteria of having at least two hydrogen bonds per base pairing. The remaining planar, doubly hydrogen-bonded triples are expected to be highly stable (12). The filtering mechanism therefore screens out less stable triple arrangements, which in large complexes such as the ribosomal subunits, are expected to be numerous but not necessarily of any interest to the user. However, users are given an option of turning off this screening mechanism to increase the search coverage. MC-annotate does not filter out its output for triples in the same way and therefore also includes base triple interactions that may have single

A

	Hits (chain) residuetype+number	Pattern/Motif information
☑	(0) G 691 , (0) A 692 , (0) A 693 , (0) A 694	GAAA tetraloop
☑	(0) G 805 , (0) A 806 , (0) A 807 , (0) A 808	GAAA tetraloop
☑	(0) G 1629 , (0) A 1630 , (0) A 1631 , (0) A 1632	GAAA tetraloop
☑	(0) G 2412 , (0) C 2413 , (0) A 2414 , (0) A 2415	GAAA tetraloop
☑	(0) G 1863 , (0) C 1864 , (0) A 1865 , (0) A 1866	GCAA tetraloop
☑	(0) G 2249 , (0) G 2250 , (0) G 2251 , (0) A 2252	GGGA tetraloop
☑	(0) G 2696 , (0) A 2697 , (0) G 2698 , (0) A 2699	GAGA tetraloop
☑	(0) G 1055 , (0) U 1056 , (0) A 1057 , (0) A 1058	GUAA tetraloop
☑	(0) G 2877 , (0) U 2878 , (0) A 2879 , (0) A 2880	GUAA tetraloop
☑	(0) G 577 , (0) C 578 , (0) G 579 , (0) A 580	GCGA tetraloop
☑	(0) U 1770 , (0) U 1771 , (0) C 1772 , (0) G 1773	UUCG tetraloop
☑	(0) G 469 , (0) U 470 , (0) G 471 , (0) A 472	GUGA tetraloop
☑	(0) G 2630 , (0) U 2631 , (0) G 2632 , (0) A 2633	GUGA tetraloop

B

|	Hits (chain) residuetype+number	Pattern/Motif information
☑	(B) A 232 , (C) C 85 , (B) G 263	A-minor Type 0 AGC
☐	(B) A 288 , (B) C 83 , (B) G 77	A-minor Type 0 AGC
☐	(B) A 96 , (B) G 2 , (B) C 342	A-minor Type 0 ACG
☐	(B) A 169 , (B) G 127 , (B) C 148	A-minor Type 0 ACG
☐	(B) A 229 , (B) G 263 , (C) C 85	A-minor Type 0 ACG
☐	(B) A 222 , (B) U 265 , (C) A 83	A-minor Type 0 AAU
☐	(B) A 48 , (B) C 319 , (B) G 46	A-minor Type I ACG
☐	(B) A 96 , (B) C 340 , (B) G 4	A-minor Type I ACG
☐	(B) A 169 , (B) C 128 , (B) G 147	A-minor Type I ACG
☐	(B) A 169 , (B) C 129 , (B) G 147	A-minor Type I ACG
☐	(B) A 191 , (B) C 73 , (B) G 87	A-minor Type I ACG
☐	(B) A 207 , (C) C 71 , (C) G 63	A-minor Type I ACG
☐	(B) A 288 , (B) C 84 , (B) G 76	A-minor Type I ACG
☐	(B) A 305 , (B) C 20 , (B) G 44	A-minor Type I ACG
☐	(B) A 207 , (C) G 63 , (C) C 72	A-minor Type I AGC
☐	(B) A 223 , (B) G 220 , (B) C 58	A-minor Type I AGC
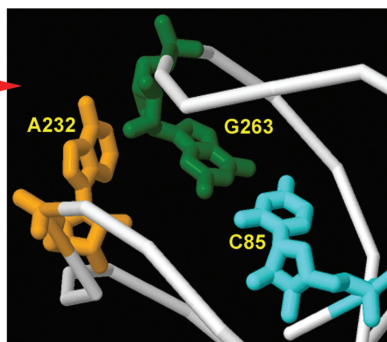
Figure 2. Snapshots of NASSAM output showing (**A**) the results of a GNRA tetraloop search for the *H. marismortui* 23S rRNA structure (PDB ID: 1FFK) in the left panel with the selected GNRA tetraloop hits viewed relative to the whole structure in a Jmol window in the right panel; (**B**) the results listing the hits to A-minor motifs in the structure of a bacterial ribonuclease P (PDB ID: 3Q1R) in the left panel and a close up examination in Jmol of one selected hit (C85.A232.G263).

hydrogen bonds connecting each base (Table 1). However, executing the NASSAM search at a distance tolerance of 60% revealed two more base triples that were not annotated by MC-annotate but which fitted the planarity and two hydrogen bond per base criteria (Table 1).

Another NASSAM versus MC-annotate comparison, using the 'annotate all patterns except base pairs' option, was performed for a group II intron structure from *Oceanobacillus iheyensis* (PDB ID: 3G78) (26,27). Both NASSAM and MC-annotate retrieved 14 base triples and A-minor motifs, but upon closer scrutiny, there were differences in the results retrieved. We found that MC-annotate yielded three triples that concurred with our filtering parameter and three A-minors, but the other eight hits did not fit the NASSAM filtering criteria (Table 2). On the other hand, the patterns retrieved by NASSAM consisted of six base triples, three of which were not retrieved by MC-annotate and eight A-minor motifs, five of which were not retrieved by MC-annotate (Table 2). Visual analysis of the triples annotated by NASSAM showed that one of the additional NASSAM triples (G267.U298.G320) overlaps with another G267.C286.G320 triple to form a base-quadruple interaction (G267.C286.U298.G320). Both MC-annotate and NASSAM annotated the same set of T-loop and tetraloop structures (data not shown), but only NASSAM was able to annotate the two ribose zipper motifs (C272.U273.G92.A93 and C6.C7.A318.A319) presumably due to its use of an updated database of motifs reported in the literature. These results further illustrate the complementarity of the NASSAM and MC-annotate programs that are a result of the different approaches used by the two programs to provide an annotation solution for RNA structures.

One caveat for the user is that, due to the nature of the base arrangements that the program looks for, and the use of higher search tolerances, a retrieved hit may not always be the desired target. As an example, one of the retrieved hits for an AGC A-minor Type 0 (A232.C85.G263) in 3Q1R when visually scrutinized does not fit the definition for that particular pattern despite the spatial arrangement matched by the search doing so (Figure 2B). In this particular example, the adenine is off plane and therefore not able to insert into the minor groove at the C285.G263 base pairing (Figure 2B). To aid visual examination and assessment of the hits in conforming to a particular defined arrangement, users have the option of selecting the hit(s) of interest and viewing them in the Jmol window.

SUMMARY

Perhaps, the most important capability of NASSAM is its ability to accept any existing complex RNA structure however large. This includes the ribosomal subunits from prokaryotes (such as PDB ID: 1FFK) (28) and eukaryotes (such as PDB ID: 3U5D) (29). This thus enables NASSAM to be used as a standard first line annotation tool for tertiary motifs not only in complex structures such as ribozymes but also in large complex assemblies such as the ribosomal subunits.

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. Nagaswamy,U., Larios-Sanz,M., Hury,J., Collins,S., Zhang,Z., Zhao,Q. and Fox,G.E. (2002) NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.*, **30**, 395–397.
3. Tinoco,I.J. (1993) In: Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 603–607.
4. Duarte,C.M. and Pyle,A.M. (1998) Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.*, **284**, 1465–1478.
5. Petrov,A.I., Zirbel,C.L. and Leontis,N.B. (2011) WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.*, **39**, W50–W55.
6. Dror,O., Nussinov,R. and Wolfson,H. (2005) ARTS: alignment of RNA tertiary structures. *Bioinformatics*, **21(Suppl. 2)**, ii47–ii53.
7. 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.
8. 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.
9. Nguyen,M.N., Tan,K.P. and Madhusudhan,M.S. (2011) CLICK—topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res.*, **39**, W24–W28.
10. Liu,Y.C., Yang,C.H., Chen,K.T., Wang,J.R., Cheng,M.L., Chung,J.C., Chiu,H.T. and Lu,C.L. (2011) R3D-BLAST: a search tool for similar RNA 3D substructures. *Nucleic Acids Res.*, **39**, W45–W49.
11. Harrison,A.M., South,D.R., Willett,P. and Artymiuk,P.J. (2003) Representation, searching and discovery of patterns of bases in complex RNA structures. *J. Comput. Aided Mol. Des.*, **17**, 537–549.
12. Firdaus-Raih,M., Harrison,A.M., Willett,P. and Artymiuk,P.J. (2011) Novel base triples in RNA structures revealed by graph theoretical searching methods. *BMC Bioinformatics*, **12**, S2.

13. Spriggs,R.V., Artymiuk,P.J. and Willett,P. (2003) Searching for patterns of amino acids in 3D protein structures. *J. Chem. Inf. Comput. Sci.*, **43**, 412–421.

14. Ullmann,J.R. (1976) An algorithm for subgraph isomorphism. *JACM*, **23**, 31–42.

15. Xin,Y. and Olson,W.K. (2009) BPS: a database of RNA base-pair structures. *Nucleic Acids Res.*, **37**, D83–D88.

16. Abu Almakarem,A.S., Petrov,A.I., Stombaugh,J., Zirbel,C.L. and Leontis,N.B. (2012) Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.*, **40**, 1407–1423.

17. Nissen,P., Ippolito,J.A., Ban,N., Moore,P.B. and Steitz,T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.

18. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.

19. Krasilnikov,A.S. and Mondragon,A. (2003) On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640–643.

20. Tamura,M. and Holbrook,S.R. (2002) Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.*, **320**, 455–474.

21. Correll,C.C. and Swinger,K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*, **9**, 355–363.

22. Cheong,C., Varani,G. and Tinoco,I. Jr (1990) Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature*, **346**, 680–682.

23. Jucker,F.M. and Pardi,A. (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry*, **34**, 14416–14427.

24. Walberer,B.J., Cheng,A.C. and Frankel,A.D. (2003) Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. *J. Mol. Biol.*, **327**, 767–780.

25. Gendron,P., Lemieux,S. and Major,F. (2001) Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.*, **308**, 919–936.

26. Toor,N., Keating,K.S., Taylor,S.D. and Pyle,A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.

27. Wang,J. (2010) Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr. D*, **66**, 988–1000.

28. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.

29. Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.