JAR3D Webserver: Scoring and aligning RNA loop sequences to known 3D motifs

James Roll[1], Craig L. Zirbel[1,*], Blake Sweeney[2], Anton I. Petrov[3] and Neocles Leontis[4]

[1]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA, [2]Department of Biology, Bowling Green State University, Bowling Green, OH 43403, USA, [3]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK and [4]Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA

ABSTRACT

Many non-coding RNAs have been identified and may function by forming 2D and 3D structures. RNA hairpin and internal loops are often represented as unstructured on secondary structure diagrams, but RNA 3D structures show that most such loops are structured by non-Watson–Crick basepairs and base stacking. Moreover, different RNA sequences can form the same RNA 3D motif. JAR3D finds possible 3D geometries for hairpin and internal loops by matching loop sequences to motif groups from the RNA 3D Motif Atlas, by exact sequence match when possible, and by probabilistic scoring and edit distance for novel sequences. The scoring gauges the ability of the sequences to form the same pattern of interactions observed in 3D structures of the motif. The JAR3D webserver at http://rna.bgsu.edu/jar3d/ takes one or many sequences of a single loop as input, or else one or many sequences of longer RNAs with multiple loops. Each sequence is scored against all current motif groups. The output shows the ten best-matching motif groups. Users can align input sequences to each of the motif groups found by JAR3D. JAR3D will be updated with every release of the RNA 3D Motif Atlas, and so its performance is expected to improve over time.

INTRODUCTION

JAR3D (Java-based Alignment of RNA using 3D structure), a recently announced software package (1) available for download at https://github.com/BGSU-RNA/JAR3D, is now available as a web server at http://rna.bgsu.edu/jar3d/. JAR3D identifies possible 3D geometries of user-provided hairpin (HL) and internal (IL) loop sequences from predicted RNA 2D structures. The JAR3D web server takes the sequence(s) of RNA HL or IL as user input and scores them against known RNA 3D motif groups from the RNA 3D Motif Atlas (2), returning possible matches. IL and HL have many important functions, including providing binding sites for proteins or small molecules, anchoring RNA tertiary interactions, and playing architectural roles such as introducing bends in helices or changing their twist. Inferring the 3D structure of internal and hairpin loops is also an important step on the way toward correctly predicting complete RNA 3D structures from sequence and potential protein–RNA and RNA–RNA interactions.

Input sequences do not have to match exactly the sequences of motif instances in the 3D Motif Atlas to produce viable hits, as JAR3D allows for inserted 'bulged' nucleotides and isosteric substitutions that preserve conserved interactions. JAR3D only returns known geometries; it does not predict new geometries that have not been observed. Importantly, the underlying RNA 3D Motif Atlas is updated as new RNA 3D structures are solved experimentally, expanding the set of possible motifs that can be matched. The relatively small number of possible geometries returned by JAR3D are amenable to downstream physics-based energy calculations or targeted mutagenesis experiments, to further narrow down the range of possible 3D geometries. The server has been running since 2012. Results from JAR3D have already been used in published work (3,4).

Other available web servers will match a single input sequence to known loops exactly (5,6) or up to a given edit distance (7). Downloadable packages range from fragment assembly methods for *de novo* prediction of full 3D geometries (8,9) to prediction of non-canonical basepairs only (10) to matching sequences to a small number of probabilistic models (11,12).

MATERIALS AND METHODS

The RNA 3D Motif Atlas extracts all RNA HL and IL from a non-redundant (NR) set of RNA 3D structures from the PDB/NDB and clusters these loops into geomet-

rically similar families, called motif groups ([2]). For each motif group in each release of the Motif Atlas, we construct a probabilistic model for sequence variability, based on a hybrid Stochastic Context-Free Grammar/Markov Random Field (SCFG/MRF) method ([1]). To parameterize each model, we use all instances of the motif in the Motif Atlas and knowledge of RNA nucleotide interactions, especially isosteric basepairs and their substitution patterns ([13,14]), which are summarized in the RNA Basepair Catalog available at http://ndbserver.rutgers.edu/ndbmodule/services/BPCatalog/bpCatalog.html. Given the sequence of an HL or of the two strands of an IL, each SCFG/MRF model calculates an *alignment score* that reflects the probability that the sequence can form a given 3D motif. From the alignment score, we calculate the *alignment score deficit*, which is the difference between the maximum alignment score over sequences from 3D instances in the motif group and the alignment score of the given sequence.

Edit distance provides another measure of the similarity between a given sequence and the 3D instances in a motif group. We calculate the Levenshtein edit distance between the interior bases (excluding flanking bases) of each strand of the input sequence and the corresponding strands of each sequence from 3D instances, then take the minimum over 3D instances, calling the result the *interior edit distance.* (For diagnostic purposes, we also display the *full edit distance,* which includes the flanking basepairs as well.)

Each motif group has an acceptance region defined in terms of the alignment score deficit and the interior edit distance, which is described using the inequality:

Alignment score deficit $+ 3 *$ Interior edit distance $\leq$ Motif-group-specific constant

The motif-group-specific constant is set to limit the rate at which certain randomly-generated sequences fall into the acceptance region; the rate is kept to 4%, which reduces JAR3D's false positive rate. To provide a reasonably uniform scoring system across motif groups, the *cutoff score* is defined to be 100 at the point where alignment score deficit and interior edit distance both equal 0, and it decreases linearly to 0 on the line at the edge of the acceptance region, where equality obtains in the formula above. The acceptance regions have been shown to be as good as or better than RMDetect, for each motif scored by RMDetect ([1,11]).

INPUT

The user can input a single HL or IL sequence or a set of multiple sequences, all of which are thought to form the same 3D motif. The two strands of each IL must be separated by the * character. Multiple input sequences need not have the same length or be aligned to each other. Alternatively, the user can input one or more full-length RNA sequences (up to 1000 nucleotides) containing multiple HL and IL, if preceded by a separate line of input with the dot-bracket notation defining the 2D structure. Users should make every effort to predict accurate 2D structures for input RNAs, using covariation data where available. Successful use of JAR3D depends on input of correct 2D structures. The user may specify the release of the RNA 3D Motif Atlas to use. The user may also provide an optional title to be

associated with the query. Sample input for different scenarios is provided on the input page.

PROCESSING

JAR3D separates sequences accompanied by a dot-bracket 2D structure annotation into individual HL and IL and processes each loop separately. For each loop, JAR3D scores all input sequences against all HL or IL models in the requested version of the Motif Atlas, calculates the acceptance rate (percentage of input sequences falling into the acceptance region), mean cutoff score and median edit distances against each motif group, and stores these in a database for later retrieval and display.

PERFORMANCE

The JAR3D webserver admits a motif group as a potential match when one or more of the input sequences fall within the acceptance region of the motif group, and ranks these motif groups from highest acceptance rate to lowest in its output. Key performance questions when submitting sequences to JAR3D are whether or not the 'correct' motif group will have a non-zero acceptance rate and whether the 'correct' motif group will be listed first, with the highest acceptance rate.

To address these questions, we considered, in turn, each 3D instance of IL from each motif group in release 1.13 of the Motif Atlas; we refer to the 3D instance as the *original instance* and its motif group as the *original motif group.* Where possible, we mapped the original instance to a multiple sequence alignment from GreenGenes ([15]) for bacterial 16S rRNA, alignments from Silva ([16]) for all other ribosomal SSU or LSU RNAs, or from alignments retrieved from a site indicated by the authors of RNASTAR ([17]) for various small RNAs ([1]). We extracted the corresponding columns from the alignments and use the rows of this *alignment extract* as a source of sequence variants of the original instance, noting that the same IL sequence may appear multiple times on different rows of the alignment extract. In some cases, two 3D instances of IL from the same motif group are mapped to the same alignment columns, as for example when two IL from bacterial rRNA are mapped to the same columns of the same alignment; then we only use the alignment extract once. To focus on *novel* sequences, we excluded sequences having the same interior sequence as a 3D instance from the original motif group. Doing so excludes virtually all of the potential novel sequence variants for very small IL, so we exclude from consideration IL having just one or two interior nucleotides; their signal to noise ratio is too low. This leaves 344 alignment extracts from 344 original instances representing 163 original motif groups.

The alignment extracts we used are imperfect for a variety of reasons including sequencing error, mis-alignment and changes in 3D geometry across the organisms in the alignment (which, when verified, strongly suggests that the sequences are not alignable). We cleaned up the alignment extracts to a minimal degree, excluding sequences with non-ACGU characters and sequences having too few nucleotides to make flanking Watson–Crick pairs on one strand or the other. Visual inspection suggests that many of
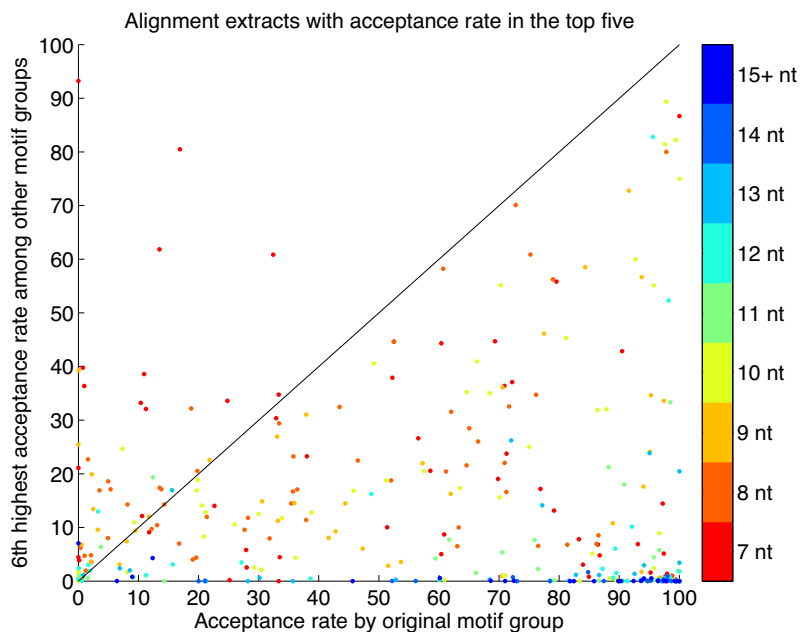
Figure 1. Each dot represents one of 344 IL alignment extracts. Horizontal position indicates acceptance rate by the original motif group, vertical position represents 6th highest acceptance rate of the same sequences by other motif groups, and color represents number of nucleotides in the original motif group. Each alignment extract below the diagonal line will have the original motif group in the top five results. Regarding the horizontal position, 80 extracts (23.3%) have acceptance rate below 20%, 46 (13.4%) have acceptance rate between 20% and 40%, 26 (7.6%) have acceptance rate between 40% and 60%, 59 (17.2%) have acceptance rate between 60% and 80% and 133 (38.7%) have acceptance rate above 80%.

the remaining sequences are unlikely to form the same 3D geometry as the original 3D instance, for instance, because of extreme differences in sequence length. A full discussion is beyond the scope of this article, but even noisy data give an indication of JAR3D's ability to recognize novel sequence variants.

Of these 344 alignment extracts, 334 or 97.1% had non-zero acceptance rate from the original motif group, with the average acceptance rate being 57.7%. The acceptance rate over the 344 alignment extracts is the horizontal variable shown in Figure 1. Nearly 40% of the alignment extracts have acceptance rate over 80%. Thus, if one inputs novel sequences of a 3D motif known to JAR3D, there is a reasonable probability that at least one of them will be accepted by the correct motif group, and the probability increases as one inputs additional sequences, as there are strictly more ways for a match to occur. This answers the first question in the affirmative.

Second, how does the acceptance rate of an alignment extract by the original motif group compare to the acceptance rate by other motif groups? In 183 alignment extracts (53.2%), the acceptance rate by the original motif group is higher than any other group, meaning that with enough input sequences, the original motif group will appear first in the JAR3D results. (Assuming the sequences are drawn with replacement from the distribution in the alignment extract, by the strong law of large numbers the observed acceptance rate against each motif group will converge to the population average, which is higher for the original motif group.) In 283 alignment extracts (82.3%), the acceptance rate for the original motif group is in the top five, and in 304 alignment extracts (88.4%), the acceptance rate for the original

motif group is in the top 10, meaning that with enough sequences from these extracts, the original motif group will appear in the top 5 or 10 JAR3D results. Figure 1 illustrates the 'top five' results in detail: each dot represents one of the 344 alignment extracts, and the horizontal position of the dot indicates the acceptance rate by the original motif group, while the vertical position of the dot indicates the 6th highest acceptance rate across other motif groups; thus dots below the black diagonal line represent alignment extracts for which the original motif group will score in the top five. The dots are colored by the number of conserved positions in the original motif group as shown by the colorbar. Note that alignment extracts from larger motifs (blue dots) tend to have high acceptance rate by the original motif group and low acceptance rates by other groups, which will make them stand out clearly in JAR3D results. On the other hand, alignment extracts from smaller motifs, especially those of 7 or 8 nucleotides (red and orange dots), will sometimes have high acceptance rates from several motif groups, making it more difficult to identify the 'correct' motif group from acceptance rate alone. Indeed, the 3D geometry of small loops is difficult to identify from sequence alone, in part because some sequences can adopt multiple 3D geometries, and so there may not be a single 'correct' motif group for a set of input sequences, but only a number of possible 3D geometries.

OUTPUT

The JAR3D results page shows up to 10 3D motif groups having the highest non-zero acceptance rates, with links to the RNA 3D Motif Atlas. Diagnostic information (DI) is also provided and includes the acceptance rate, the cutoff

score and minimum edit distances to sequences of known instances of each 3D motif. (In selecting the top 10 motif groups, ties in acceptance rate are broken by mean cutoff score.) Alignments of input sequences to motif groups are displayed on supplementary output pages. JAR3D result pages are persistent URLs (which can be re-visited later) with long randomly-generated strings.
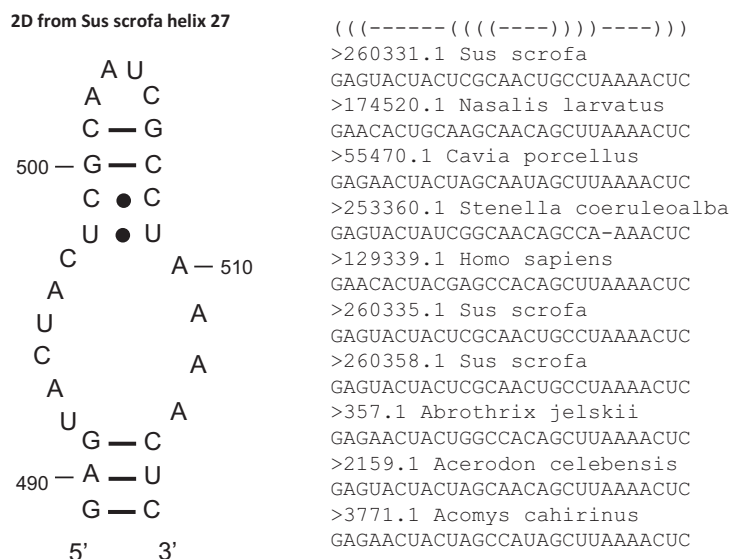
ILLUSTRATIVE EXAMPLES

Mitochondrial rRNA provides interesting examples of the capabilities and potential uses of the JAR3D server. Multiple sequence alignments of the mammalian mitochondrial (mmt) 12S ribosomal RNA (rRNA) are available from the comparative RNA website (CRW) (18) and can be retrieved using R3D-2-MSA (19). We submitted these sequences to the JAR3D web server using Release 1.18 of the RNA 3D Motif Atlas dating from December 2014, posted well before the first high resolution mmt small subunit (SSU) ribosome structure was solved. Thus, we can test the ability of JAR3D to identify mitochondrial SSU rRNA loop sequences with probabilistic models constructed without the mitochondrial rRNA 3D data, and use the solved structure to evaluate the matches from JAR3D.

We focus on helix 27 of the SSU rRNA, a conserved helical element that contains one IL and one HL, both in mitochondria and in bacteria. The left panel of Figure 2 shows the predicted 2D structure for helix 27 of Sus scrofa from CRW. Note that the IL and HL are displayed as loops having no internal structure; this is a typical starting point for an investigation of loop structure with JAR3D. An alignment of 308 mammalian helix 27 sequences was obtained from R3D-2-MSA using the link http://rna.bgsu.edu/r3d-2-msa?units=5AJ3|1|A||489:5AJ3|1|A||516&aid=2 (19). (One sequence, from 55471.1 Cavia porcellus, was omitted because it has non-ACGU characters.) The right panel of Figure 2 shows 10 of

the submitted sequences of the alignment accompanied by a manually produced dot-bracket secondary structure line.

The alignment of the remaining 307 sequences of helix 27, along with the dot-bracket secondary structure shown in Figure 2, was submitted to the JAR3D input page to obtain possible 3D structures for the loops. We gave the query the custom title, 'Mitochondria, SSU, helix 27, mammals only.' The result page can be seen at the persistent link http://rna.bgsu.edu/jar3d/result/a223a5a5-67f9-4011-be6c-60be02d50134/. A condensed version of the result page is shown in Figure 3, which we now describe from top to bottom. The result page first shows the randomly-assigned query ID. The associated URL permits users to view results can be accessed again in the future, as they are saved in our database indefinitely. The next line shows the version of the Motif Atlas used to process the query, in this case version 1.18. The custom title is shown next, followed by a summary of the user's input. If the user's input includes a secondary structure, column numbers are added to the input summary alignment; each column number is read vertically from top to bottom.

Below the input summary, each submitted IL and HL is listed separately, ordered by the lowest column number included in the loop. The input sequence(s) of each loop are shown in a text box, and the column numbers for each loop are provided to assist in finding loop locations in the secondary structure and alignment. Next, up to 10 potential matches are shown, ranked first by the acceptance rate and then by mean cutoff score. (In rare cases, the user will be interested in more than 10 results; a link is provided to see all results for that loop.) The best possible cutoff score is 100, and a negative cutoff score indicates rejection by the motif group. More information on JAR3D models and their acceptance criteria can be found in the JAR3D paper (1). Columns labeled 'Full Edit Distance' and 'Interior Edit Distance' on the results page show the median (over

2D from Sus scrofa helix 27

```
(((------(((----))))----)))
>260331.1 Sus scrofa
GAGUACUACUCGCAACUGCCUAAAACUC
>174520.1 Nasalis larvatus
GAACACUGCAAGCAACAGCUUAAAACUC
>55470.1 Cavia porcellus
GAGAACUACUAGCAAUAGCUUAAAACUC
>253360.1 Stenella coeruleoalba
GAGUACUAUCGGCAACAGCCA-AAACUC
>129339.1 Homo sapiens
GAACACUACGAGCCACAGCUUAAAACUC
>260335.1 Sus scrofa
GAGUACUACUCGCAACUGCCUAAAACUC
>260358.1 Sus scrofa
GAGUACUACUCGCAACUGCCUAAAACUC
>357.1 Abrothrix jelskii
GAGAACUACUGGCCACAGCUUAAAACUC
>2159.1 Acerodon celebensis
GAGUACUACUAGCAACAGCUUAAAACUC
>3771.1 Acomys cahirinus
GAGAACUACUAGCCAUAGCUUAAAACUC
```

Figure 2. The left panel shows the predicted secondary structure of helix 27 of the mitochondrial small ribosomal subunit with sequence from Sus scrofa. The full 2D structure is available from the CRW at the URL http://www.rna.ccbb.utexas.edu/RNA/Structures/x.16.m.S.scrofa.5AJ3.pdf The right panel shows the dot-bracket secondary structure from the predicted 2D followed by 10 sequences from the alignment extract produced by R3D-2-MSA.

JAR3D Query a223a5a5-67f9-4011-be6c-60be02d50134 completed

Run on Motif Atlas Version 1.18

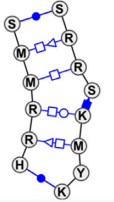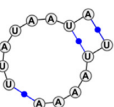## Mitochondria, SSU, helix 27, mammals only

Input Summary

```
0000000001111111111222222222
12345678901234567890123456789
============================
(((------((((----))))----)))
>260331.1 Sus scrofa
GAGUACUACUCGCAACUGCCUAAAACUC
>174520.1 Nasalis larvatus
GAACACUGCAAGCAACAGCUUAAAACUC
```

Loop 1

Column Numbers: 3-10, 21-26

```
GUACUACU*UAAAAC
```

Click on headings to reorder table - **View All Results for this Loop**

| #	Matching motif	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	**Motif IL_95652.6** Basepair signature: cWW-tSH-tHH-cSH-tWH-tHS-L-cWW [Align sequences]	99.02	35.28	3	2	
2	**Motif IL_73656.1** Basepair signature: cWW-L-R-R-R-L-L-L-cWW-cWW [Align sequences]	22.15	-11.84	6	4	

Loop 2

Column Numbers: 13-18

```
CAACUG
```

Click on headings to reorder table - **View All Results for this Loop**

#	Matching motif	Acceptance Rate	Mean Cutoff Score	Full Edit Distance	Interior Edit Distance	2D Diagram
1	**Motif HL_48116.3** Basepair signature: cWW-s35-s55-s35-s55-s35 [Align sequences]	94.14	55.67	1	1	
2	**Motif HL_19626.4** Basepair signature: cWW-s35-cSW-s35-s35 [Align sequences]	93.81	30.74	2	1	

Figure 3. Condensed JAR3D result page for helix 27 query. Details of the query are given at the top, followed by the best two motif matches for the IL and two for the HL.
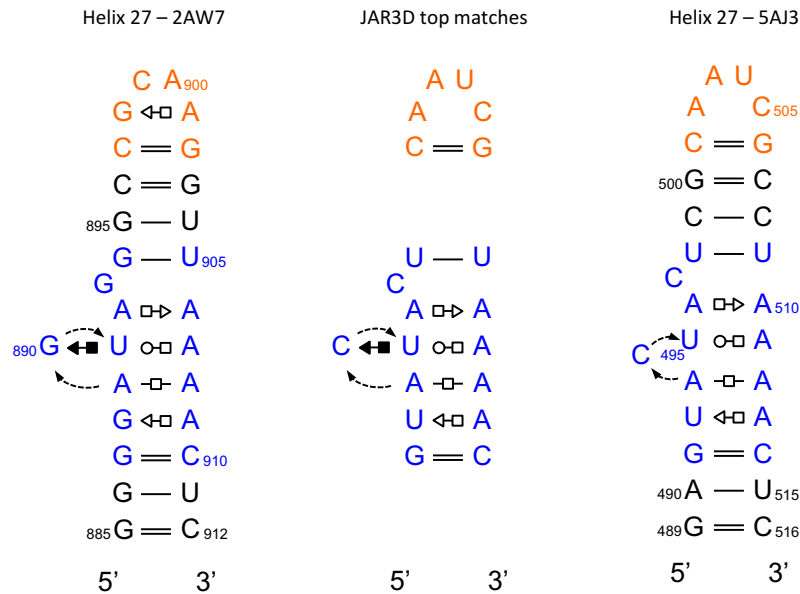
Figure 4. The left panel shows the annotation of the 3D structure of helix 27 from *Escherichia coli* 16S rRNA (PDB id 2AW7), which is highly conserved in other bacteria. The center panel shows annotations of the best IL and HL matches to mmt helix 27 sequences from JAR3D, with sequence from Sus scrofa, following the alignment to the motif group, cf. Figure 5. The right panel shows an annotation of the 3D structure of helix 27 of the Sus scrofa mitochondrial SSU (PDB id 5AJ3).

all input sequences) of the minimum (over 3D sequences) interior and full edit distances from the input sequences to the sequences of 3D structures in the motif group. Full edit distance includes all nucleotides in the motif, interior edit distance excludes the flanking Watson-Crick basepairs. Finally, a 2D VARNA (20) basepair diagram for the motif group is shown, as is the *basepair signature,* which is a text version of the list of consensus basepair interactions.

Results for internal loop of h27

For Loop 1 of the output shown in Figure 3, we see that 99.02% of the input sequences fall in the acceptance region of motif group IL_95652.6, with a mean cutoff score of 35.28. This acceptance rate and mean cutoff score are very good for such a large number of sequences. The next best match, IL_73656.1, only accepts 22.15% of input sequences and has a negative mean cutoff score. Thus, only the match to IL_95652.6 warrants further investigation. One can learn more about motif group IL_95652.6 by following the link in the left column to its page in the 3D Motif Atlas. One can view an alignment of the input sequences to motif group IL_95652.6 by clicking the Align Sequences button in that row. JAR3D alignments are discussed further below.

We can evaluate the best JAR3D IL match (center panel of Figure 4, blue) by comparing to the homologous position in a bacterial 3D structure (left panel of Figure 4) and by looking at the 3D structure of Sus scrofa mitochondrial small ribosomal subunit (right panel of Figure 4). By homology, we would predict that the IL of mammalian mitochondrial helix 27 should have the basepairs observed in bacterial 16S rRNA as shown in the left panel of Figure 4, and the JAR3D match in the center panel agrees in all respects. However, the actual 3D from Sus scrofa shows that C494 is bulged out of the motif and instead interacts with

a mitochondrial specific SSU protein. It does not make the expected base triple observed in bacterial 16S rRNA. The alignment to the motif group shown in Figure 5 confirms that G almost never occurs in this position in the mmt helix 27 sequences. Thus, JAR3D correctly predicts most aspects of the geometry of this IL, but nature has settled on a sequence substitution that alters the geometry in this one respect.

Results for hairpin loop of h27

In bacteria, the HL in helix 27 makes a GNRA motif; cf. left panel of Figure 4, orange, and by homology we may expect that mmt helix 27 would also have a GNRA loop. However, the JAR3D output does not list a GNRA motif group as a likely structure. Instances in the top-matching motif group, HL_48116.3, actually contain no intra-loop basepairs, and loop bases instead make tertiary interactions. The right panel of Figure 4 shows that bases of Sus scrofa helix 27 also make no intra-loop basepairs and the overall structure differs from the GNRA HL observed in bacteria. Thus, the matches proposed by JAR3D reflect reality better than homology does. We note that predicting long-range tertiary interactions is beyond the scope of JAR3D.

JAR3D alignments to motif groups

By clicking 'Align sequences' on the result page, users can display alignments of input sequences to a selected motif group returned by JAR3D. These alignments include the sequences from the 3D instances that make up the motif group, to allow users to identify locations of insertions or deletions, as well as base substitutions implicated in crucial stabilizing interactions.

JAR3D Query a223a5a5-67f9-4011-be6c-60be02d50134

Alignment to IL_95652.6

Run on Motif Atlas Version 1.18

Input sequences are initially listed on the top, with a grey background. Sequences from 3D structures are initially listed on the bottom, with a white background.
The strands of your internal loop have been switched to match the strand ordering seen in IL_95652.6. They are still in 5' to 3' order.

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Meets Cutoff	Cutoff Score	Interior Edit Distance	Full Edit Distance
Position		1		2		3		4		5		6				7		8		9		10		11		12		13		14					
Insertion	I		I		I		I		I		I		I		I		I		I		I		I		I		I		I		I				
>260331.1 Sus scrofa		U		A		A		A		A		C		*		G		U		A		C		U		A		C		U		True	32.194866	2	3
>174520.1 Nasalis larvatus		U		A		A		A		A		C		*		A		C		A		C		U		G		C		A		False	-9.237416	3	5
>55470.1 Cavia porcellus		U		A		A		A		A		C		*		G		A		A		C		U		A		C		U		True	32.348810	2	3
>253360.1 Stenella coeruleoalba				A		A		A		A		C		*		G		U		A		C		U		A	U	C				False	-52.314964	3	5
>129339.1 Homo sapiens		U		A		A		A		A		C		*		A		C		A		C		U		A		C		G		True	35.071026	2	3
IL_1FJG_034		U		G		A		A		A		C				G		G		A		G		U		A		C		G					
IL_2AW7_041		U		A		A		A		A		C				G		G		A		G		U		A		C		G					
IL_3J7A_050		A		G		A		A		A		G				C		G		A		G		U		A		U		U					
IL_3U5F_057		U		G		A		A		A		C				G		G		A		G		U		A		U		G					
IL_4BPP_052		U		G		A		A		A		C				G		A		A		G		U		A		U		G					
IL_4CUX_055		U		G		A		A		A		C				G		G		A		G		U		A		U		G					
IL_4W23_083		C		G		G		C		C		C				G		A		G		C		G		C		U		G					

Figure 5. Alignment of mammalian mitochondrial rRNA helix 27 IL sequences to IL_95652.6. The alignment uses the strand ordering of IL_95652.6. Five the sequences from the helix 27 alignment are at the top, followed by the 7 sequences from 3D structures used to form IL_95652.6.

Figure 5 shows a section of the alignment of the mammalian mitochondrial helix 27 IL sequences to IL_95652.6. Figure 5 has been condensed to show rows that are of particular interest. The full alignment can be seen at the persistent link http://rna.bgsu.edu/jar3d/result/a223a5a5-67f9-4011-be6c-60be02d50134/0/IL_95652.6. User input sequences are shown first highlighted with a grey background, and below that sequences from 3D structures are shown with a white background. The mitochondrial sequences align well, but with the bulged G at position 10 replaced by C. The order of sequences can be changed by clicking on any column header, to reorder the table based on the contents of that column. To the right of the alignment are columns indicated the sequences accepted by the model, their cutoff scores and the minimum interior and full edit distances to sequences from 3D. Sequences not likely to form the 3D structure of the motif group are easily identified in the output.

CONCLUSION

The JAR3D web server can be used to investigate possible 3D geometries for one or more sequences of an RNA HL or IL identified in a proposed 2D structure of an RNA molecule. A variety of input formats are supported. The result page shows up to 10 matches with links to known 3D geometries. Alignment pages allow the user to compare input sequences to known 3D motif groups and 3D sequences. As the RNA 3D Motif Atlas is updated on an ongoing basis with new loop instances from new RNA 3D structures, existing motif groups will grow and new ones will be created, allowing JAR3D to match a wider range of input sequences to known RNA geometries.

In contrast to the standalone JAR3D program (1), the webserver accepts full RNA sequences (with a secondary structure) and separates out individual HL and IL as a convenience to the user, it sorts motif groups by decreasing acceptance rate of the input sequences, it provides 2D and 3D views of matching motif groups plus basepair signatures, it provides an easy-to-access visualization of the alignment of input sequences to a given motif group, all of these displays are interactive and provide ready links to the RNA 3D Motif Atlas, and all are available through persistent URLs which will be maintained indefinitely.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

Conflict of interest statement. None declared.

REFERENCES

1. Zirbel,C.L., Roll,J., Sweeney,B.A., Petrov,A.I., Pirrung,M. and Leontis,N.B. (2015) Identifying novel sequence variants of RNA 3D motifs. *Nucleic Acids Res.*, **43**, 7504–7520.
2. Petrov,A.I., Zirbel,C.L. and Leontis,N.B. (2013) Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA*, **19**, 1327–1340.
3. Kladwang,W., Chou,F.C. and Das,R. (2012) Automated RNA structure prediction uncovers a kink-turn linker in double glycine riboswitches. *J. Am. Chem. Soc.*, **134**, 1404–1407.
4. Akkuratov,E.E., Walters,L., Saha-Mandal,A., Khandekar,S., Crawford,E., Zirbel,C.L., Leisner,S., Prakash,A., Fedorova,L. and Fedorov,A. (2014) Bioinformatics analysis of plant orthologous introns: identification of an intronic tRNA-like sequence. *Gene*, **548**, 81–90.
5. Chojnowski,G., Walen,T. and Bujnicki,J.M. (2014) RNA Bricks–a database of RNA 3D motifs and their interactions. *Nucleic Acids Res.*, **42**, D123–D131.
6. 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.
7. Schudoma,C., May,P., Nikiforova,V. and Walther,D. (2010) Sequence-structure relationships in RNA loops: establishing the basis for loop homology modeling. *Nucleic Acids Res.*, **38**, 970–980.
8. Das,R., Karanicolas,J. and Baker,D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.
9. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
10. Honer zu Siederdissen,C., Bernhart,S.H., Stadler,P.F. and Hofacker,I.L. (2011) A folding algorithm for extended RNA secondary structures. *Bioinformatics*, **27**, i129–i136.
11. Cruz,J.A. and Westhof,E. (2011) Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat. Methods*, **8**, 513–521.
12. Theis,C., Honer Zu Siederdissen,C., Hofacker,I.L. and Gorodkin,J. (2013) Automated identification of RNA 3D modules with discriminative power in RNA structural alignments. *Nucleic Acids Res.*, **41**, 9999–10009.
13. Stombaugh,J., Zirbel,C.L., Westhof,E. and Leontis,N.B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.*, **37**, 2294–2312.
14. Sweeney,B.A., Roy,P. and Leontis,N.B. (2015) An introduction to recurrent nucleotide interactions in RNA. *Wiley Interdiscip. Rev. RNA*, **6**, 17–45.
15. McDonald,D., Price,M.N., Goodrich,J., Nawrocki,E.P., DeSantis,T.Z., Probst,A., Andersen,G.L., Knight,R. and Hugenholtz,P. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.*, **6**, 610–618.
16. Quast,C., Pruesse,E., Yilmaz,P., Gerken,J., Schweer,T., Yarza,P., Peplies,J. and Glockner,F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, **41**, D590–D596.
17. Widmann,J., Stombaugh,J., McDonald,D., Chocholousova,J., Gardner,P., Iyer,M.K., Liu,Z., Lozupone,C.A., Quinn,J., Smit,S. *et al.* (2012) RNASTAR: an RNA STructural Alignment Repository that provides insight into the evolution of natural and artificial RNAs. *RNA*, **18**, 1319–1327.
18. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
19. Cannone,J.J., Sweeney,B.A., Petrov,A.I., Gutell,R.R., Zirbel,C.L. and Leontis,N. (2015) R3D-2-MSA: the RNA 3D structure-to-multiple sequence alignment server. *Nucleic Acids Res.*, **43**, W15–W23.
20. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.