Collecting and Characterizing RNA Molecules with Known Secondary Structure

Mario Enriquez, Zsuzsanna Hollander, Jessica Zheng Zhao
Dept. of Computer Science, The University of British Columbia
enriquez, zsuzsa, zzhao@cs.ubc.ca

March 15, 2004

Abstract

RNAs are important elements in the metabolism of all organisms. The function of many RNAs is crucially dependent on their structure. Secondary structure is one of the most important factors determining 3d-structure, and functions. Having a more in-depth knowledge about the secondary structure of RNAs would provide researchers a better insight for understanding their functionality. In the last two decades, numerous approaches have been developed that try to predict the secondary structure of an RNA sequence and pairs of RNA molecules.
The work presented in this paper is an attempt to provide researchers with a better understanding of the secondary structure of various RNA types and possibly improve the methods for prediction of secondary structures. We assembled a database containing RNA molecules with their known secondary structures and integrated it with an existent RNA Structure Analyzer. The integration of the aforementioned tools provides structural statistics for the molecules in the database and also characterizes the properties of various types of biological RNA structures.

1 Introduction

RNAs are important elements in the metabolism of all organisms [1]. The function of many RNAs is crucially dependent on their structures [2]. Having a more in-depth knowledge about the secondary structure of RNAs would provide researchers a better insight for understanding their functionality. In the last two decades, numerous approaches have been developed that try to predict the secondary structure from an RNA sequence [3, 4, 5, 6] and pairs of RNA molecules [7]. However, the prediction accuracy of these programs needs to be further refined, in particular for long and complex molecules.

One motivation for the work presented in this paper is to help researchers get a better understanding of the secondary structure of various RNA types and possibly improve the methods for

prediction of secondary structures. Another motivation stems from the problem of RNA secondary structure design. The design of RNA molecules with specific structural properties has many potential applications [2]. Having additional knowledge about the common properties of certain types of RNA structures may ultimately assist researchers to improve current design methods to help create RNA molecules that are a close match to naturally occurring ones. The goal of this work is to assemble a database containing RNA molecules with their known secondary structures and integrate it with an existent RNA Structure Analyzer. The integration of the aforementioned tools will provide structural statistics for the molecules in the database and also characterize the properties of various types of biological RNA structures. There exists a number of RNA databases [8, 9, 10, 11, 15, 16, 23, 25, 30] available on the internet out of which some have the secondary structure included and some require the use of an additional tool, like MC-Annotate [17] to get the secondary structure. There are RNA analysis tools, as well, [4, 14, 33, 34] but as far as we know there are no systems that include a collection of RNA molecules and analysis tools to get statistical information about the structure of the RNAs.

The first stage of this work addresses the creation of a database of single RNA molecules and pairs of RNA molecules with their known (or predicted) secondary structure. We collected data from existing databases (see Data Collection section for details), and also looked up and manually collected RNA structures from diverse publications, such as Nucleic Acids Research, The Journal of Biological Chemistry, Science, Nature, and others, with the help of the PubMed Library. The information regarding each RNA molecule contained in the database will be introduced in Section 2 Implementation.

Our database stores RNA structures in bpseq format. In this format, each base in the molecule is assigned a position and a pairing with another base in the molecule if the base is paired (see figure 1). In this database, each molecule (or pair of molecules) is stored in an individual file. For the case of pairs of RNA molecules, we use a similar format as for single RNA molecules. Each base in both sequences is assigned a unique and consecutive position id so the second RNA molecule is recorded right after the end of the first. The range of positions for each RNA molecule is stated clearly in the comment field of the file.

For the second stage of our work, we extended an existing RNA Analyser (implemented by Farheen Rawji) and integrated it with our database system. The original Analyser has several limitations:
- handles only pseudoknot-free RNA sequences,
- can not provide comprehensive statistics for a set of RNA molecules but can only analyze a single RNA sequence,
- is operated via a text-based menu and command line without a graphical user interface.

Since pseudoknot formations are common in many kinds of RNA molecules and are responsible for several important RNA functions [13] [14], we integrated the RNA Analyser with our system in such way that the system is able to deal with pseudoknotted structures. We believe that having a comprehensive summary of the common properties for specific types of RNAs is useful for both RNA structure prediction and RNA sequence design. If, for example, a researcher is designing an

Position	Base	Pointer
1	G	20
2	U	19
3	C	18
4	U	17
5	G	16
6	A	15
7	A	14
8	C	13
9	G	0
10	A	0
11	A	0
12	A	0
13	G	8
14	U	7
15	U	6
16	C	5
17	G	4
18	G	3
19	A	2
20	U	1

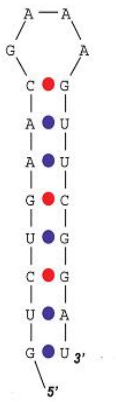

Figure 1: The left side table shows a sample molecule represented in bpseq format. The right side figure shows the molecule's graphical representation.

rRNA sequence, our database can provide information regarding common stem lengths, loop sizes, and other information from existing rRNA sequences to serve as guidelines for design. Therefore, the system developed for this work also provides a fully functional database system which allows the user to interactively query RNA sequences satisfying certain constraints, such as sequence length, type, etc., via an easy-to-use graphical user interface. Additionally, statistical reports of structural properties can be generated for any type of RNA specified by the user. Furthermore, a visualization tool - representing RNA secondary structures in a graphical form - is available for users to view a single RNA sequence or to do pair-wise structure comparisons.

Currently existing RNA databases mostly contain predicted secondary structures. Most of the secondary structures for real RNA sequences had to be taken from publications and manually captured into our database. Furthermore, to the best of our knowledge, existing databases fail to provide summary statistics for the structures' basic characteristics. This is one of the main advantages of the work presented here.

In the remainder of this paper, we explain the functionality of our system in further detail in Section 2 and present the obtained statistical results from the sequences and structures captured in our database in Section 3. Our conclusions and proposed future work are presented in Section 4.

2 Implementation: Database Design and Functionality

Our database system runs under the Microsoft Windows platform (95, 98, 98se, NT, ME, 2000, XP). The code was written in Microsoft Visual Basic 6.0. The code interfaces with the original RNA Analyser using command line arguments and text files. The database system was implemented using text files in .csv format which can be edited/viewed using any spreadsheet application. Each RNA sequence is stored in an individual file. These files are stored in comma-separated-value format. This means that every field in the database is separated by a comma. This allows easy editing and viewing of these files in any spreadsheet software. Additionally, the database stores essential information in an index which is loaded to memory for efficient searching. This index contains the basic information about each molecule (everything except the sequence itself). We require sequential access to this index, since we allow queries with different combinations of constraints for the different fields in the database. The information contained in the index for each molecule is presented in table 1. Each molecule occupies 5988 bytes (about 6k bytes) in this index. This means that for a database containing 1000 molecules, the index size is 6Mb. We load this index to memory since accessing information in memory is considerably faster than that stored in a hard drive. This index is loaded into memory during the system's start up to facilitate browsing and searching the database.

Some of the single RNA molecules and pairs of RNA molecules collected had secondary structures with base triplets; a base being paired with two or more other bases. These sequences, although not analyzable with our database system, were nonetheless included in the database under a special folder. The existence of base triplets in the sequence is mentioned in the Comments field in the Sequence Information (top middle part of figure 2). The way these base triplets were described in bpseq format is as follows:

Position	Base	Pointer
32	G	7, 26

Here we can see that base G at position 32 is paired with bases at positions 7 and 26.

Interface
Our database system's main window is presented in figure 2. Within this screen the user can submit queries, edit RNA information and visualize sequence secondary structures. Additionally, this screen presents users with the option of obtaining statistical data for sequences that meet certain query criteria.

The functionalities are illustrated as follows:

- **Query**
  The user can query any of the RNA sequence-structure files by filling in the fields (RNA type, sequence length, organism, etc.) to be queried in the "Query" form and clicking the

Data	Explanation
Filename	Stores the Filename for the Sequence
Organism	Oraganism to which the molecule belongs
Accesion Number	Accesion Number
SequenceType	Sequence Type (rRNA, snRNA,etc.)
Citation	Website
Source	Source for molecule (database,website)
RelatedPublications	Publication where molecule was found
Length of Sequence	Number of bases in sequence
Sequence In a String	Contains the sequence with no structure information.
Comments 1	Comments
Comments 2	Comments
Containing Folder	Database Folder where file is located
Experimentally Proven T/F	specifies wether the structure is verified.
Number of Stems	Total number of stems in the molecule
Number of Hairpins	Total number of Hairpins in the molecule
Number of Bulges	Total number of Bulges in the molecule
Number of Interior Loops	Number of interior loops in molecule
Number of Multi Loops	Number of multi-loops in molecule
Maximum number of Stems per Multi Loop	Maximum number of stems in multi-loops
Minimum number of Stems per Multi Loop	Minimum number of stems in multi-loops
Average number of Stems per Multi Loop	Average number of stems in multi-loops
Total number of stems for External Loop	Total number of stems in external loop
Average stem length	Average stem length in molecule
Average Hairpin Size	Average Hairpin size in molecule
Average Bulge Size	Average Bulge size for molecule
Number of Pseudoknotted pairs	Number of pseudoknotted pairs

Table 1: Information for molecules contained in database index file.

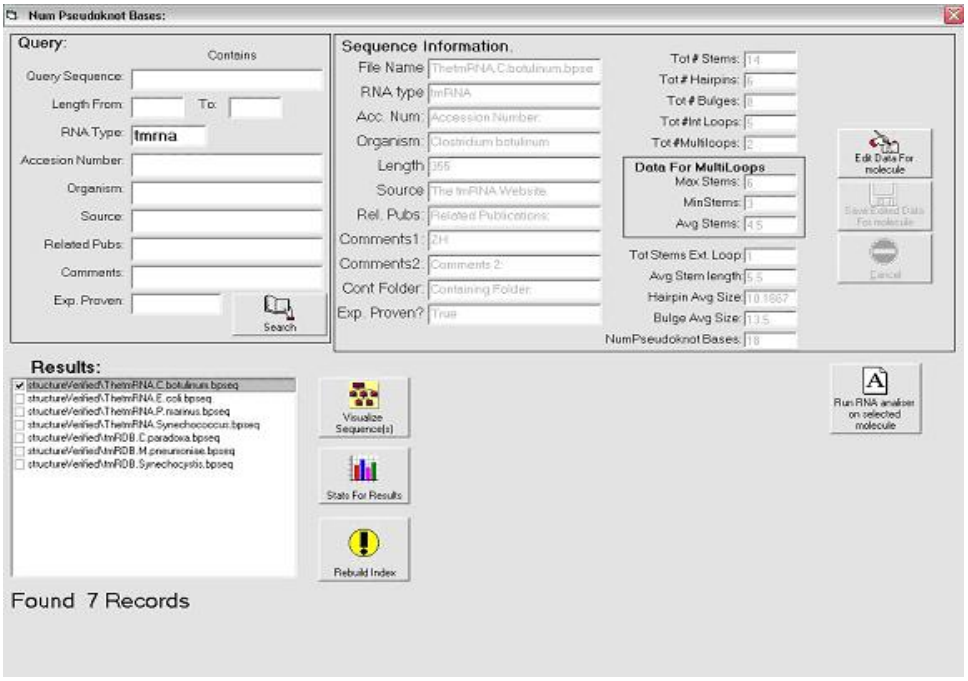

Figure 2: Database System Main Screen

”Search” button. Those fields left blank, are ignored by the query engine. The results for the query will be returned in the ”Results” window. From this results window, a user can select a sequence, to view the detailed information about this particular RNA. This information will be displayed in the ”Sequence Information” window. The results present the data obtained from the RNA file in the top middle part of the screen. The right part of the screen presents the results obtained from the RNA Analyser as further explained within the Structural Analysis functionality.

- **Edit**
  The user can modify a file stored in the database. After a sequence has been selected from the results window, the user can edit the information contained therein by clicking the ”Edit data for molecule” button and changing the desired data fields. This modified information is stored to the file and index when the user presses the ”Save edited data for molecule” button. At any point during this process, the user can opt to cancel the operation by clicking ”cancel”.

- **Visualization**
  The user can select one or two RNA files from those listed in the ”Results” window to visualize their secondary structures. By clicking the ”Visualize Sequences” button, the RNA secondary structure will be displayed as a linked graph. Figure 3 shows an example of such a graph for a single molecule. The red links in the displayed graph represent pseudoknots contained within the RNA sequence; these pairs are the minimal number of bases to be removed to make the structure pseudoknot-free. The user can also select two RNA sequences to make a visual comparison of their structure via our database system as shown in Figure

4.This graph represents bases and base pairings by connecting the bases with lines. The top half of the figure presents one molecule, the bottom half the other molecule. The red lines represent pseudoknots.

Figure 3: Visualization of a single molecule.

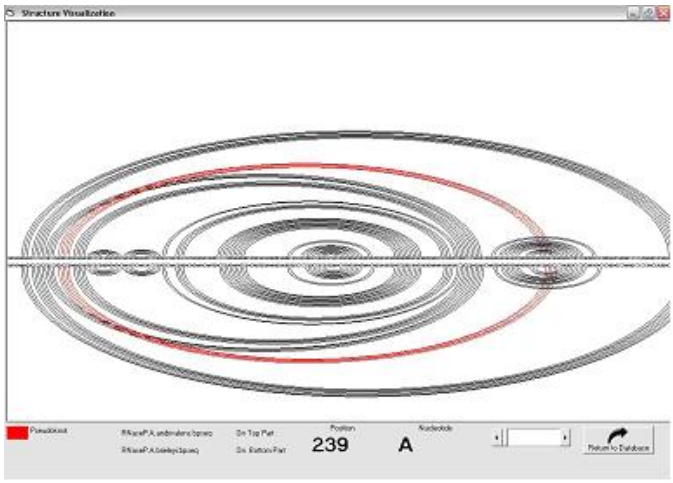

Figure 4: Visualization of two molecules.

- **Structural Analysis**
  The structural analysis for the molecule is performed in part by using ”RNA Analyser”. This analysis tool utilizes dot-bracket notation to represent secondary structure. Pseudoknots can not be represented using dot-bracket notation, thus our database must provide a pseudoknot-free structure to this analysis tool to use it for structural analysis. Our pseudoknot elimination algorithm guarantees that the minimum number of base pairs will be removed to attain a pseudoknot-free structure. Our algorithm searches through the base pairs in the molecules and identifies which pairings generate pseudoknots. The algorithm then searches for the

Statistic
Total number of stems/molecule
Total number of hairpins/molecule
Total number of bulges/molecule
Total number of interior loops/molecule
Total number of multiloops / molecule
Total number of multiloops / molecule
Maximum stems per multiloop
Minimum stems per multiloop
Average stems per multiloop
Total stems in exterior loop
Average stem length
Average hairpin size
Average bulge size
Number of pseudoknotted bases

Table 2: Statistics obtained for every molecule analyzed.

minimum number of pairings to be removed to obtain a pseudoknot-free molecule structure. The time complexity of our algorithm is O(n^2) and its operation is based on removing the base pairs with the most base-pair-crossings (see figures 3 and 4). Table 2 shows the results of the analysis performed on each molecule along with an explanation of each term used.

The detailed analysis of each type of secondary structure component (stem, hairpin loop, bulge, multi-loop, external loop) is stored in the database and presented on the upper right part of the main screen (figure 2) when a molecule is selected from the results window after a query. A detailed account of this analysis containing statistics of each type of component can be obtained by pressing the ”Run RNA Analyser on selected molecule” button. The output is displayed in a text window. The following is a sample output:

DETAILED STRUCTURAL ANALYSIS:

Stems:
S1:
Length: 4 (base pairs)
Position: 31-34, 48-51.
Connected to: S2, H1, B1.
Multiloops this stem is connected to: None
. . .
Hairpins:
H1
Length: 13 (unpaired bases)
Position: 35-47.
Type: not a special hairpin .
Connected to: S2.
. . .

Bulges:
B1
Length: 2 (unpaired bases)
Position: 52-53.
Connected to: S1, S2.
Multiloops this bulge is connected to: None
...
Interior Loops:
IL1
Length: 4, 3 (unpaired bases)
Position: 24-27, 57-59.
Connected to: S4, S2.
...

Multiloops:
ML1:
Length: 5 1 3 (unpaired bases)
Number of stems: 3
Position: 10-16, 68-70, 106-110
Connected to: S4, S6, S7
...
External Loop:
5'unpaired bases 2 .
3'unpaired bases 1 .
EL
Length: 2
Number of stems: 96
Position: 1-2, 117-120 , 234-237
Connected to: S7, S14, S21, S28, S3
...
Pseudoknot:
Pair: 26-82, ..
...

- **Summarized Statistics**
  After the user chooses a group of RNAs satisfying some query constraints, a summarized statistics report can be generated by clicking "Stats for Results" button. A secondary window will be shown. This window contains a global statistic report (see figure 5) for each type of structural component in a graphical format. The user can click each graph to enlarge it to full screen size. Each bar in the graph represents a value for a single RNA structure file. Once the graph is enlarged (figure 8), the user can select a) to sort the data presented therein (see Figure 6 and 7) compare the results in the graph with the respective sequence length for each data point (see figure 6). We believe this function to be useful in allowing researchers to better understand the distribution of values for the analysis results (avg. length of stems, bulges, etc.) for the quired sequences. This can allow a researcher to get an idea of the general layout of a certain type or types of RNA.

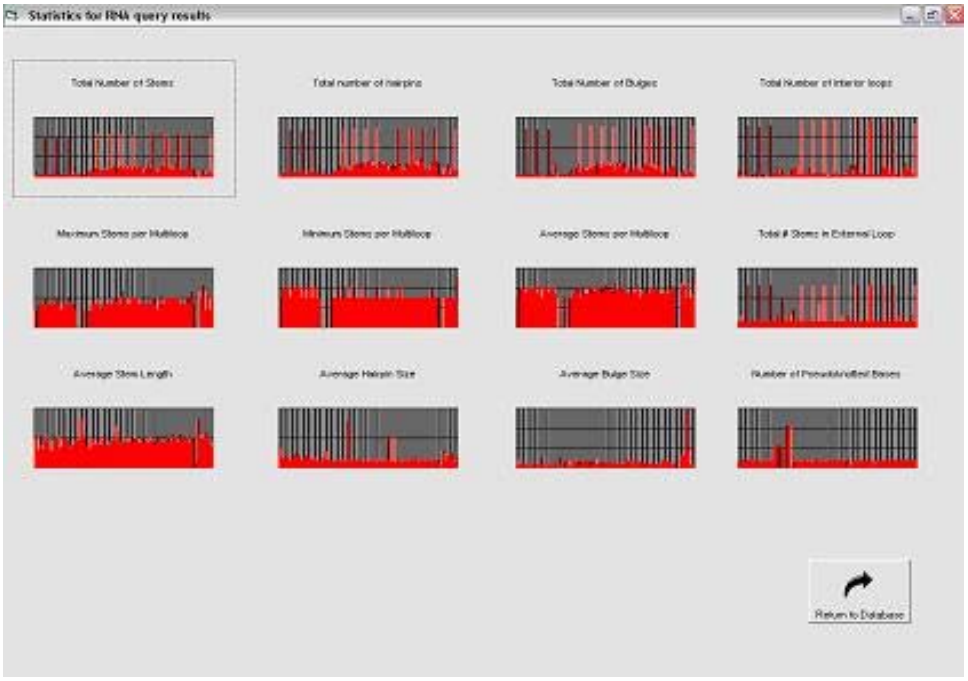

Figure 5: Global statistic report. This figure shows the statistics obtained for a series of molecules resulting from a query performed by a user. Each graph shows the results for a different statistic. In every graph, each data point represents a value for a single molecule.

3 Data Collection and Results

Data Collection
Many single RNA molecule databases exist, yet most of them contain either RNA sequences without secondary structure information or the secondary structure is presented in a diagram format. For those databases that do have structure information, many of them contain predicted structures instead of experimentally verified ones. We found a very small number of RNA sequences with known structure from existing databases that satisfied our requirements. Within these databases, data is often found in other formats rather than bpseq, which required further translation to maintain consistency with our database. Currently, our database consists of 511 rRNA sequences with predicted structures and 137 sequences whose structures are verified. The following is a summary list:

124 Single RNA with known secondary structure
- 66 RNase P RNAs [11]
- 25 tRNAs [18, 30]
- 10 ribozymes [16, 19, 20, 21, 22]
- 7 tmRNAs [10, 23]
- 7 rRNAs [16, 24, 25, 30]
- 4 snRNAs [26, 27]
- 2 snoRNAs [28, 29]

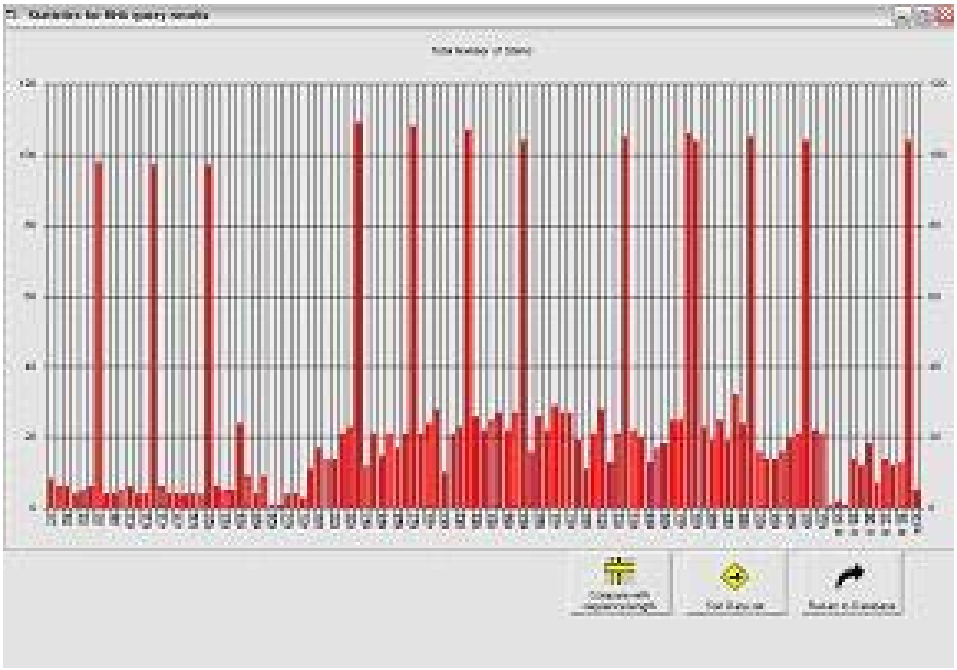

Figure 6: Detailed visualization of a single statistic for the molecules within the query results. Each data point represents the value for the statistic for a single molecule in the query.

- 1 mRNA [30]
- 1 pre-mRNAs [15]
- 1 gRNAs [31]

511 single RNA with predicted secondary structure
- 511 rRNAs [8]

13 RNA pairs with known secondary structure
- 5 tmRNA [10]
- 3 rRNAs [15]
- 2 tRNAs [15]
- 2 snoRNAs [35]
- 1 mRNA and gRNA [32]

Among these 137 verified molecules, 50 were collected from published research papers. Many of these research papers give RNA sequence-structure information only in a diagram which had to be translated manually. For pairs of RNA molecules, the secondary structure is hard to find, since no such database is available. Except for those sequences found in the Protein Data Bank, the pairs of RNA molecules were found scattered in diverse journal publications. These papers provide only a diagram with the secondary structure and had to be converted into bpseq format manually. We

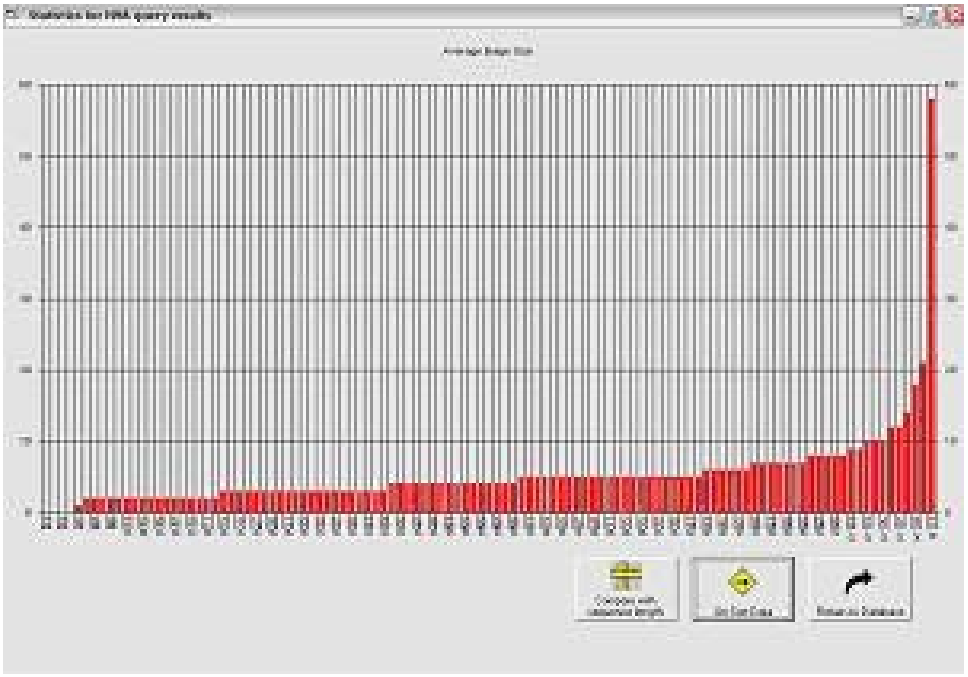

Figure 7: Sorted visualization of single statistic for the resulting molecules. Each data point represents the value for the statistic for a single molecule in the query.

provide all useful resources in Appendix A and illustrate major challenges in Appendix B.

Software Implementation

Our idea was to make this database system highly usable and intuitive. This required several iterations of design and testing by our research group. The implementation of the visualization tool and the pseudoknot finding routines required not only careful design but a series of trials to ensure proper functioning. Additionally, we intended to allow further extension of both the database contents as well as our system. These constraints forced us to utilize a certain format for the DB files to facilitate this purpose.

Statistical Results

As pointed out in the Summarized Statistics Section, we obtained a detailed analysis for every molecule in 14 statistical measures. In this section we present a series of graphs containing some of the relevant information we found during our analysis. Figures 7 and 8 show the results obtained for the analysis of 271 rRNA molecules whose secondary structure is predicted (not verified). Figure 7 presents the average stem length statistic for the molecules. Figure 8 shows the number of pseudoknotted bases statistic.

Figures 9 and 10 show the results obtained for the analysis of 22 tRNA molecules whose secondary structure has been experimentally verified. Figure 9 presents the average stem length statistic for

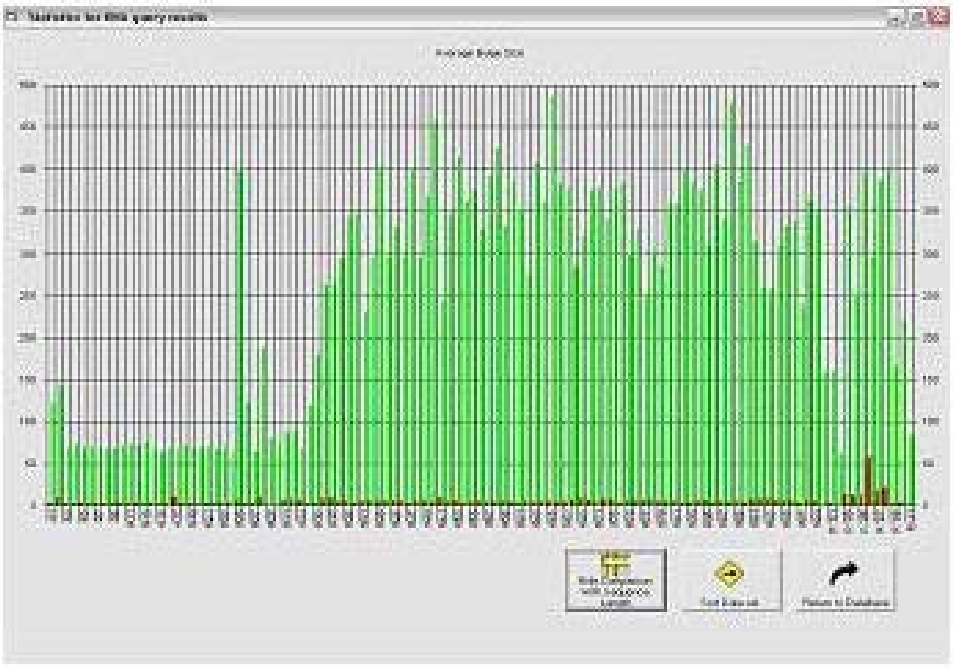

Figure 8: This figure shows a graph visualizing the results for a single statistic when compared with the sequence length for each molecule in the query results. This is useful for deciding if a certain statistic is dependent on the sequence length.

the molecules. Figure 10 shows the average hairpin size statistic. The graph showing the number of pseudoknotted bases was omitted since tRNAs have no Pseudoknots.

Figures 11 and 12 show the results obtained for the analysis of 66 RNase P RNA molecules whose secondary structure has been experimentally verified. Figure 11 presents the average stem length statistic for the molecules. Figure 12 shows number of pseudoknotted bases statistic.

Acknowledgements. We gratefully acknowledge Dr. Holger Hoos, and Mirela Andronescu, Farheen Rawji, Nan Zhang and Joseph Ajay Pathakamuri for their help and support to our project.

4 Conclusions and Future Work

In this work, we have developed an easy-to-use, accessible database of RNA sequences and their known secondary structures. In addition to providing access to this database, our system allows molecules to be compared with the aid of our structure Visualization-Comparison tool. With the help of this database RNA researchers could use our analyzer to make statistical analysis on the sequences and derive useful knowledge about the properties of various types of RNA molecules.

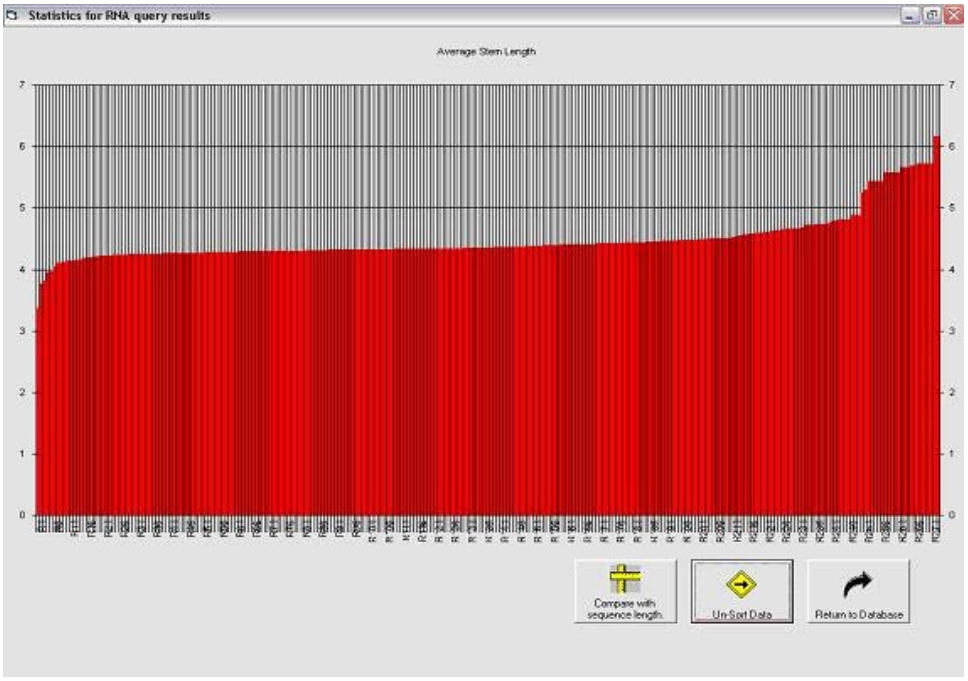

Figure 9: Average stem length results for 271 rRNA molecules.

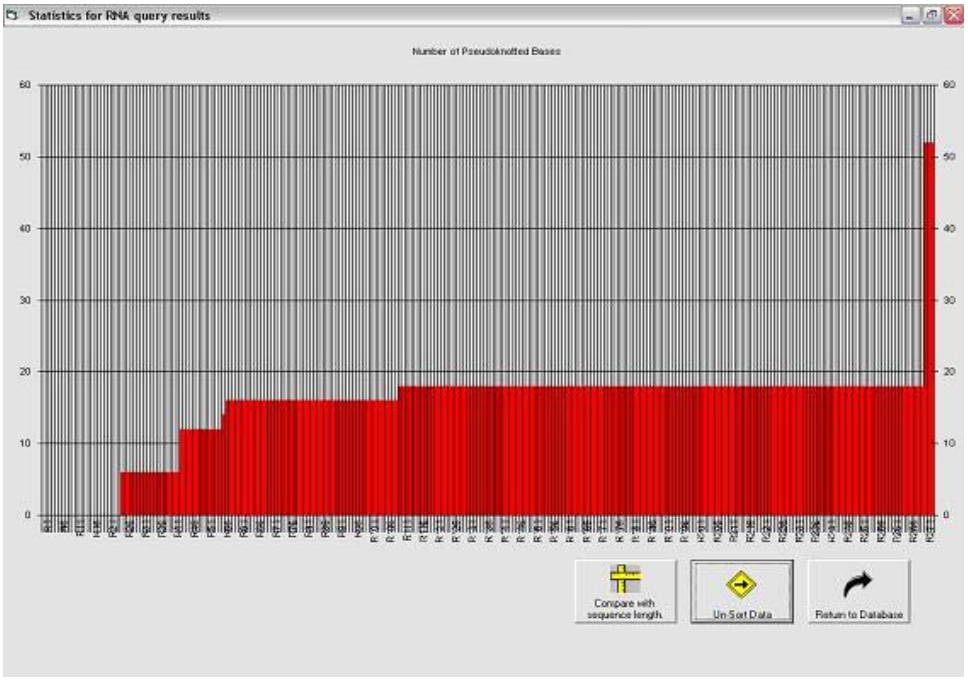

Figure 10: Number of pseudoknotted bases for the 271 predicted rRNA molecule structures.

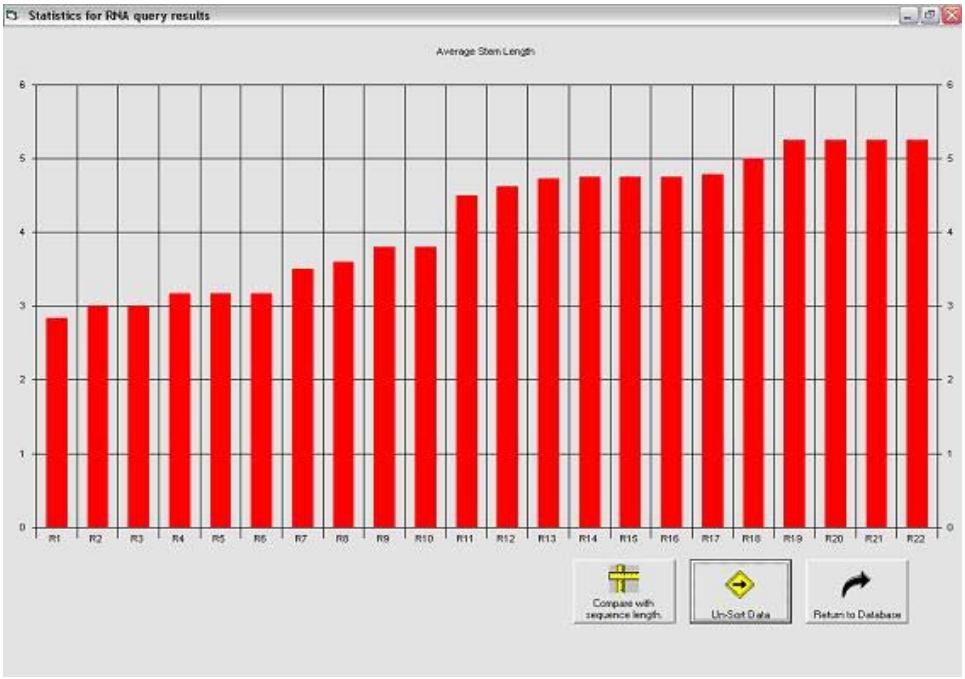

Figure 11: Average Stem Length results for 22 rRNA molecules.

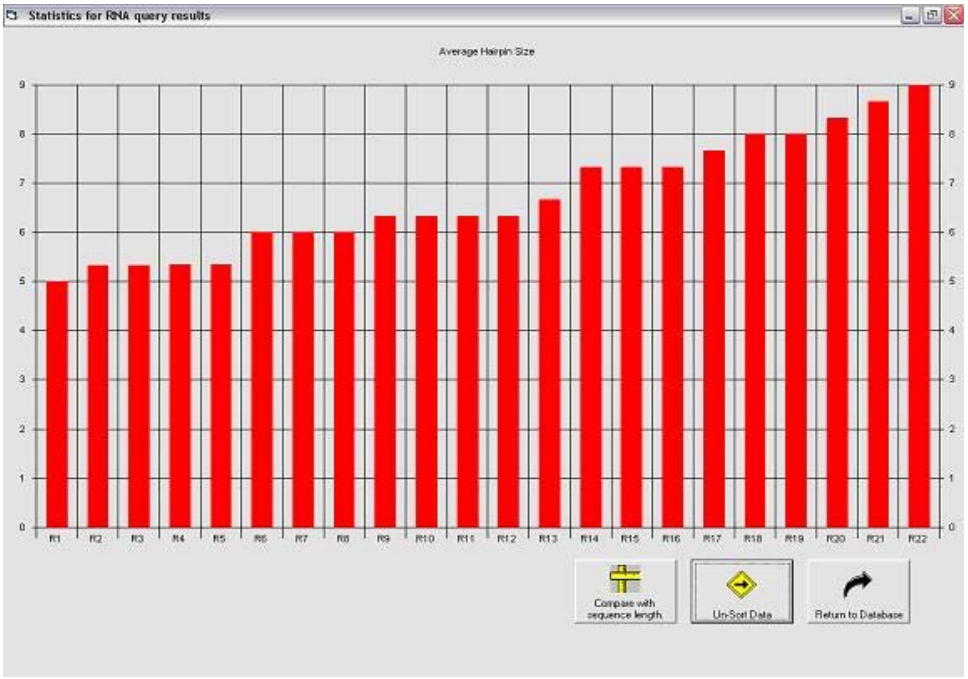

Figure 12: Average Hairpin Size results for 22 rRNA molecules.

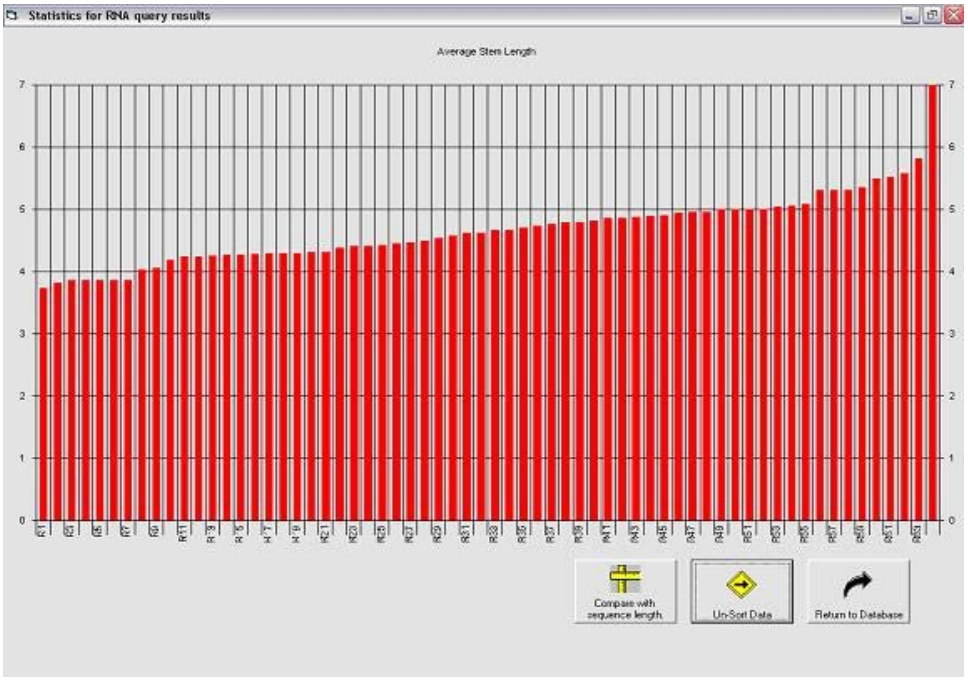

Figure 13: Average Stem Length results for 66 RNase P RNA molecules.

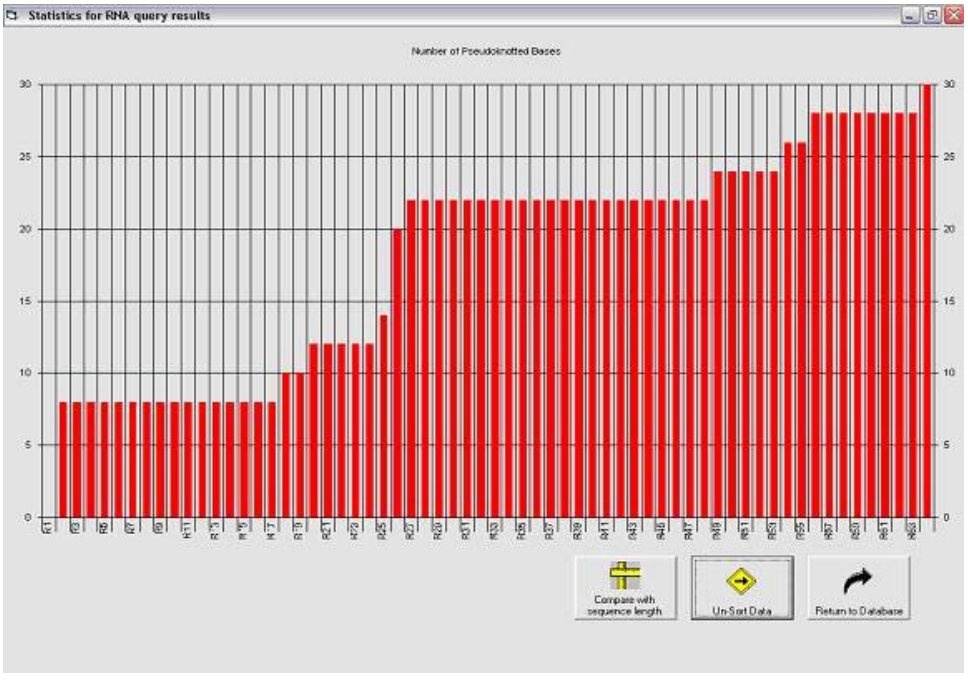

Figure 14: Number of pseudoknotted bases for the 66 verified RNase P RNA molecule structures.

Our analysis yielded a series of very useful statistics to aid in RNA design and categorization. This statistics allow researchers to compare RNA designs or predicted structures with known existing secondary structures on our database in order to validate their results or use as a guideline in design. There are some interesting directions of future improvement based on our database system which we are unable to fulfill due to the time and resource constraints. First, improve the current Analyser to be able to analyze pairs of RNA molecules. In our database, there are some pairs of RNA molecules stored. However, the current Analyser can not deal with them so the data can not be fully utilized. Second, add functionality of pair-wise sequence alignment in the visualization tool. One possible solution is to align the molecules first and then visualize the secondary structure to provide a more realistic comparison.

References

[1] M Vuyisich ”RNA Recognition by Proteins and Small Molecules: Characterization and Control of PKR-RNA Interactions and Control Protein Activity with Ligand-Regulated Aptamers (LIRAs).” 2002. http://www.chem.utah.edu/chemistry /faculty/beal/group/chapter1.pdf

[2] M Andronescu, A Fejes, et al. ”A New Algorithm for RNA Secondary Structure Design.” http://www.cs.ubc.ca/ hoos/tmp/rna-design.pdf

[3] M Zuker. ”Mfold web server for nucleic acid folding and hybridization prediction.” Nucleic Acids Res. 2003; 31(13), 3406-15.

[4] IL Hofacker, ”Vienna RNA secondary structure server”. Nucleic Acids Research, 2003; 31(13): 3429-31. Vienna RNA Package: www.tbi.univie.ac.at/ ivo/RNA

[5] GeneBee - Molecular Biology Server web page: http://www.genebee.msu.su/services/rna2_reduced.html

[6] MPS Brown. ”RNA Modeling Using Stochastic Context Free Grammars” PhD Thesis 1999. RNACAD web page: http://www.cse.ucsc.edu/ mpbrown/rnacad/

[7] M Andronescu, R Aguirre-Hernandez, A Condon, HH Hoos. ”RNAsoft: A suite of RNA secondary structure prediction and design software tools.” Nucleic Acids Res. 2003; 31(13): 3416-22. (RNAsoft website: www.rnasoft.ca)

[8] JJ Cannone, S Subramanian, MN Schnare, et al. ”The Comparative RNA Web CRW Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and other RNAs.” BMC Bioinformatics. 2002; 31: 2. $GutellLabComparativeRNAWebSite : www.rna.icmb.utexas.edu$

[9] M Sprinzl, C Horn, M Brown, A. Ioudovitch and S. Steinberg, ”Compilation of tRNA Sequences and Sequences of tRNA Genes.” Nucleic Acids. Research. (1998) 26: 148-153.

[10] KP Williams. ”The tmRNA Website: invasion by an intron.” Nucleic Acids Res. 2002; 30: 179-82. $ThetmRNAwebsite : http : //www.indiana.edu/\ tmrna/$

[11] JW Brown ”The Ribonuclease P Database.” Nucleic Acids Res. 1999; 27(1): 314. (The RNase P Database: http://jwbrown.mbio.ncsu.edu/RNaseP/home.html)

[12] PB Rupert, AR Ferre-D'Amare. ”Crystal Structure of a Hairpin Ribozyme - inhibitor complex with implications for catalysis.” Nature. 2001; 410(6830): 780-6.

[13] K Han and Y Byun, ”PSEUDOVIEWER2: Visualization of RNA Pseudoknots of any Type.” Nucleic Acids Research. 2003; 31(13): 3432-40.

[14] A Condon, B Davy, B Rastegari, et al. ”Understanding RNA Pseudoknotted Structures.” Submitted to RECOMB 2004.

[15] HM Berman, T Battistuz, TN Bhat, et al. ”The Protein Data Bank” Acta Cryst. 2002; D58: 899-907. (Protein Data Bank: http://www.rcsb.org/pdb/)

[16] HM Berman, WK Olson, DL Beveridge, et al. ”The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids.” Biophys. J., 1992; 63: 751-9. (Nucleic Acid Database: http://ndbserver.rutgers.edu/atlas/)

[17] MC-Annotate, The RNA structure evaluator. http://www-lbit.iro.umontreal.ca/mcannotate-simple/

[18] R Zardoya and A Meyer. ”Complete mitochondrial genome suggests diapsid affinities of turtles.” PNAS. 1998; 95(24): 14226-31.

[19] MJ Fedor ”Structure and Function of the Hairpin Ribozyme.” J. Mol. Biol. 2000; 297(2): 269-91.

[20] L Sun, Z Cui, RL Gottlieb, et al. ”A Selected Ribozyme Catalyzing Diverse Dipeptide Synthesis.” Chemistry & Biology. 2002; 9: 619-28.

[21] DA Lafontaine, DG Norman, and DMJ Lilley. ”Structure, folding, and activity of the VS ribozyme: importance of the 2-3-6 helical junction.” The EMBO Journal 2001; 20(6): 1415-24.

[22] EM Mobley and T Pan. ”Design and isolation of ribozyme-substrate pairs using RNase P-based ribozymes containing altered substrate binding sites.” Nucleic Acids Res. 1999; 27(21): 4298-304.

[23] C Zwieb, J Gorodkin, B Knudsen, et al. ”tmRDB (tmRNA database).” Nucleic Acids Res. 2003; 31(1): 446-7.

[24] RD Page. ”Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using manimum weighted matching.” Nucleic Acids Res. 2000; 28(20): 3839-45.

[25] M Szymanski, MZ Barciszewska, VA Erdmann, et al. ”5S Ribosomal RNA Database.” Nucleic Acids Res. 2002; 30(1):176-8.

[26] HD Cho, K Tomita, T Suzuki, et al. ”U2 Small Nuclear RNA is a substrate for the CCA-adding Enzyme.” J Biol Chem. 2002; 277(5): 3447-55.

[27] M Bell and A Bindereif. ”Cloning and mutational analysis of the Leptomonas seymouri U5 snRNA gene: function of the Sm site in core RNP formation and nuclear localization.” Nucleic Acids Res. 1999; 27(20): 3986-94.

[28] MN Schnare, JC Collings, DF Spencer, et al. ”The 28S-18S rDNA intergenic spacer from Crithidia fasciculata: repeated sequences, length heterogeneity, putative processing sites and potential interactions between U3 small nucleolar RNA and the ribosomal RNA precursor.” Nucleic Acids Res. 2000; 28 (18): 3452-61.

[29] W Speckmann, A Narayanan, R Terns, et al. ”Nuclear Retention Elements of U3 Small Nucleolar RNA.” Mol Cell Biol. 1999; 19(12): 8412-21.

[30] FH van Batenburg, AP Gultyaev, CW Pleij, et al. ”PseudoBase: a database with RNA pseudoknots.” Nucleic Acids Res. 2000; 28(1): 201-4. (PseudoBase: http://wwwbio.leidenuniv.nl/ Batenburg/PKB.html)

[31] T Hermann, B Schmid, H Heumann, et al. ”A three-dimensional working model for a guide RNA from Trypanosoma brucei.” Nucleic Acids Res. 1997; 25(12): 2311-8.

[32] SS Leung and HJ Koslowsky. ”Interactions of mRNA and gRNA involved in trypanosome mitochondrial RNA editing: Structure probing of an mRNA bound to its cognate gRNA.” RNA. 2001; 7:1803-16.

[33] P De Rijk, J Wuyts and R De Wachter ”RnaViz2: an improved representation of RNA secondary structure.” Bioinformatics. 2003; 19(2): 299-300.

[34] F Chetouani, P Monesti, P Thbault, et al. ”ESSA: An integrated and interactive computer tool for analysing RNA secondary structure.” Nucleic Acids Res 1997; 25(17): 3514-22.

[35] M Antal, A Mougin, M Kis, et al. ”Molecular characterization10at the RNA and gene levels of U3 snoRNA from a unicellular green alga, Chlamydomonas reinhardtii.” Nucleic Acids Res. 2000; 28(15): 2959-68.