



RNA SSTRAND

A new database for RNA secondary structure data
and statistical analysis of RNA structural motifs

Mirela Andronescu, Vera Bereg,
Holger H. Hoos and Anne Condon

Department of Computer Science
University of British Columbia



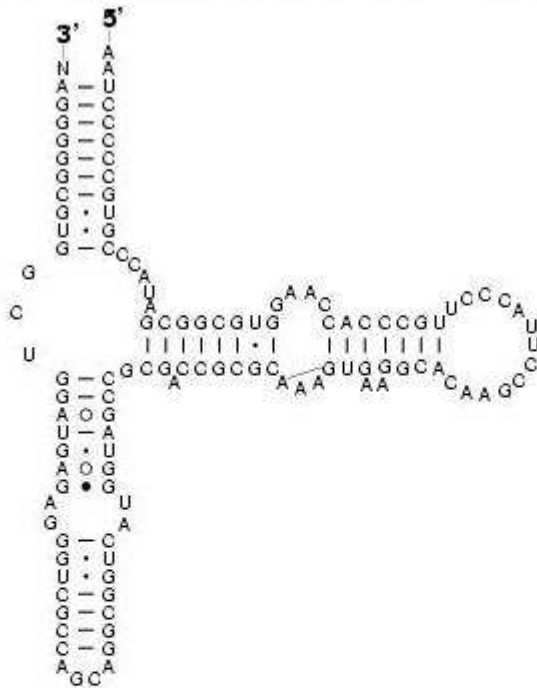
Motivation and Goals

- ◆ to interactively analyse a diverse collection of over 9000 RNA molecules
 - ◆ to offer comprehensive information on structural features within and across functional classes of molecules
 - ◆ to help improve methods for computational prediction of RNA secondary structure
 - ◆ to permit comparison across different classes of molecules
 - ◆ to allow submission of new molecules, so as to maintain the vast broadness of the database
-
-

Analyser

- ◆ in the heart of the RNA SSTRAND database lies Analyser
- ◆ it is a tool capable of taking in a bpseq representation of a secondary structure of RNA molecule and outputting statistical data that comprises the data base entries. See Figures 1, 7 and 8 for more details.
- ◆ Analyser makes use of two different algorithms: “Classifying RNA pseudo-knotted Structures” [2] and “New clique and independent set algorithms for circle graphs” [1]

Secondary structure: 5S Ribosomal RNA



Thermus thermophilus
(V01415)

1. cellular organisms 2. Bacteria
3. Thermus/Deinococcus group
4. Thermus group 5. Thermus
November 2000

Citation and related information available at <http://www.rna.icmb.utexas.edu>

Figure 1.

A graphical representation of a secondary structure of an RNA molecule.

Functionality

- ◆ RNA SSTRAND meets the outlined goals in the “Motivation and Goals” section
 - ◆ database is easily maintained through MySQL
 - ◆ access to the database is available to anyone on line at www.rnasoft.ca/sstrand
 - ◆ most of the on line content is dynamically updated upon access through MySQL queries embedded in PHP
 - ◆ simple and detailed searches through the database, as well as, statistical analysis for single and groups of RNA molecules are provided
-
-

Web Pages Provided:

- ◆ main page:
 - quick facts about the database
 - search by RNA SSTRAND ID
 - ◆ search page:
 - simple and more advanced search forms
 - ◆ analysis page:
 - analysis of several groups of RNA molecules
 - analysis of a single RNA molecule
 - ◆ submission page:
 - form for submission of an RNA molecule
 - ◆ help page:
 - definitions for all of the fields used in searches/analysis/submission forms
 - ◆ about page:
 - disclaimer
 - general statistics about the data in the database
-
-

RNA SSTRAND Search page - Mozilla

RNA SSTRAND

[[Home](#) | [Search](#) | [Analysis](#) | [Submit structures](#) | [Help](#) | [About](#)]

Search for secondary structure molecules with certain general features

| | |
|---------------------------------------|------------------|
| Type [?]: | Any type |
| Organism [?]: | |
| Source [?]: | Any source |
| | Source ID [?]: |
| Experimentally validated [?]: | I don't care |
| Number of molecules in complex [?]: | Any number and |
| Length [?]: | Any length and |

Number/sizes of secondary structure elements

| | |
|----------------------------|---------|
| Number of stems : | Any and |
| Number of A bases in stems | Any and |
| Average size of stems : | Any and |

Figure 2.

View of the search page. The first frame provides a basic search. The second frame lets specify a more detailed query.

RNA SSTRAND Search page - Mozilla

Search results: 10 results found

Search criteria:

Type: tmRNA - Transfer Messenger RNA
 Organism: -----
 Source: -----
 Source ID: -----
 Experimentally validated: I don't care
 Number of molecules in complex: Any number
 Length: Any length

| Molecule ID [?] | Molecule name [?]: | Type [?]: | Organism [?]: | Source [?]: | Source ID [?]: | Experimentally validated [?]: | Number of molecules in complex [?]: | Length [?]: |
|-------------------|--------------------------|--------------------------------|-----------------|---------------|--------------------------|---------------------------------|---------------------------------------|---------------|
| 1001 | C.botulinum tmRNA intron | tmRNA - Transfer Messenger RNA | C. botulinum | Other Sources | C.botulinum tmRNA intron | N/A | 1 | 304 |
| 1003 | E.coli tmRNA | tmRNA - Transfer Messenger RNA | E. coli | Other Sources | E.coli tmRNA | N/A | 1 | 363 |
| | Cyanophora | tmRNA - | Cyanophora | | | | | |

Done

Figure 3. A sample of search results.

RNA SSTRAND analysis page - Mozilla

RNA SSTRAND

[[Home](#) | [Search](#) | [Analysis](#) | [Submit structures](#) | [Help](#) | [About](#)]

Analyse a group of RNA molecules:

All fields are optional.

| | |
|--------------------|------------------------------------|
| RNA type | All molecules in RNA SSTRAND |
| Structural feature | Number of domains per molecule |
| Plot | Distribution of structural feature |

Perform analysis Clear form

Analyse a single RNA molecule:

The fields marked with * are mandatory.

| | |
|-------------------|----------------------|
| * RNA SSTRAND ID: | <input type="text"/> |
|-------------------|----------------------|

Perform analysis Clear form

For questions, comments, suggestions and bug reports, please contact ssstrand@rnasoft.ca
Copyright © 2004 BETA LAB - [University of British Columbia](#)

Figure 4.

View of analysis page. It is possible to search for a single molecule via its RNA SSTRAND ID or view statistics for a group of molecules.

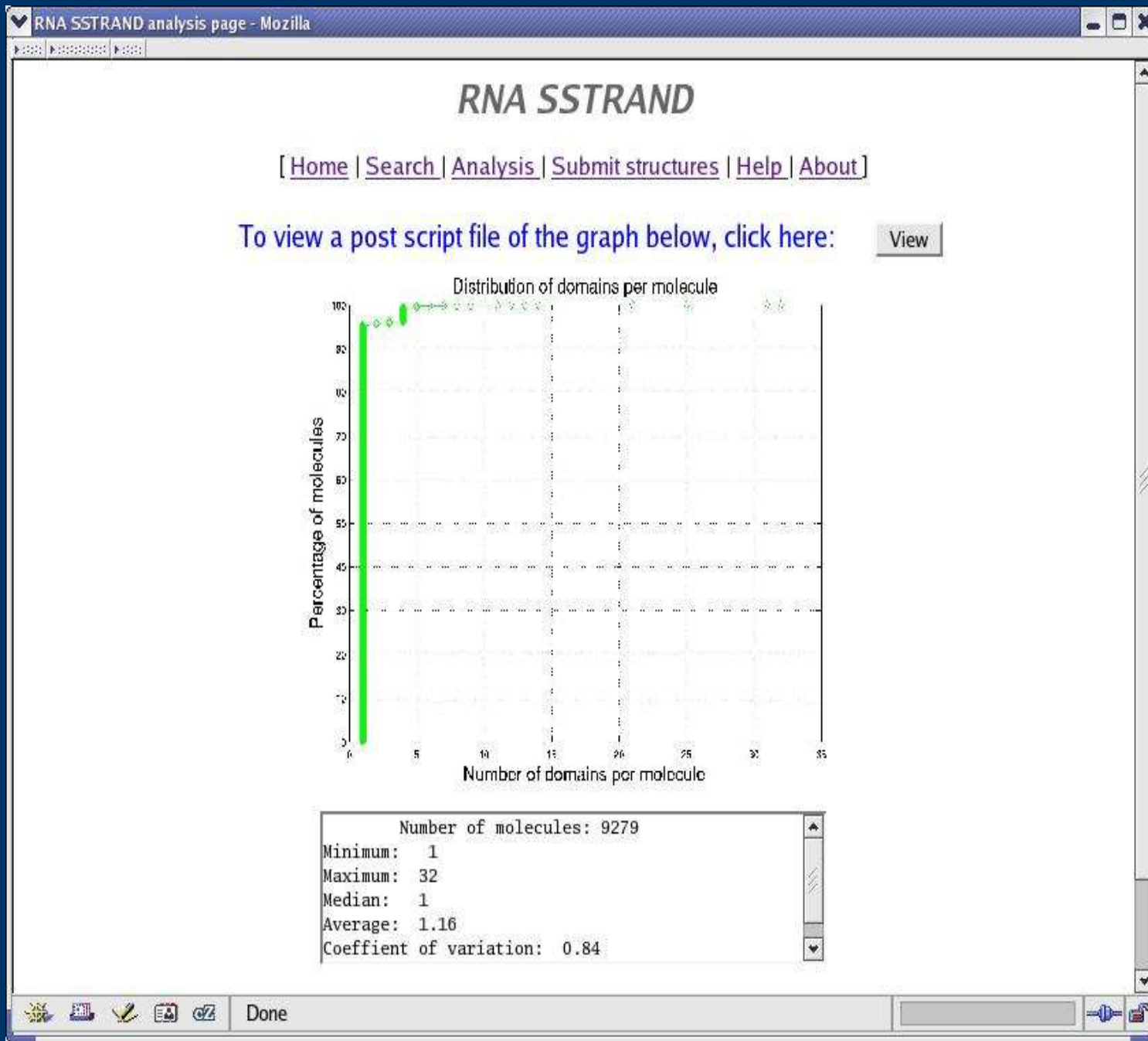


Figure 5.
A sample of analysis query result for a group of RNA molecules.

Single Molecule Web Page

- ◆ each molecule together with its quick statistics has its own web page
- ◆ this page can be accessed
 - through clicking on a id button that was shown during a search output
 - via typing in the molecule id either at the main page or analysis page prompts
 - saving the URL of the page upon first viewing and using it to come back to this page
- ◆ there are also two buttons provided for viewing of the input and output files of the analyser

RNA SSTRAND molecule page - Mozilla

RNA SSTRAND

[[Home](#) | [Search](#) | [Analysis](#) | [Submit structures](#) | [Help](#) | [About](#)]

The basic statistics for molecule 1001.

| | |
|--|--|
| Molecule ID [?]: | 1001 |
| Molecule name [?]: | C.botulinum tmRNA intron |
| Type [?]: | Transcription-Related RNA |
| Organism [?]: | C. botulinum |
| Source [?]: | Other Sources |
| Source ID [?]: | C.botulinum tmRNA intron |
| Reference [?]: | http://wilab.inha.ac.kr/pseudoviewer/ , Nucleic Acids Res. 30 (1): 179-182, 2002 |
| Experimentally validated [?]: | N/A |
| Method for secondary structure determination [?]: | N/A |
| Number of molecules [?]: | 1 |
| Fragments used [?]: | N/A |
| Length [?]: | 304 |
| Number of stems [?]: | 16 |
| Number of hairpin loops [?]: | 8 |
| Number of bulges [?]: | 2 |
| Number of internal loops [?]: | 5 |
| Number of multi loops [?]: | 1 |
| Number of pseudoknots [?]: | 1 |
| Number of bands [?]: | 2 |
| Number of unpaired bases [?]: | 104 |
| Number of paired bases [?]: | 200 |
| Number of paired bases to be broken to be pseudoknot free [?]: | 5 |
| Average stem length [?]: | 5.19 |
| Average hairpin loop length [?]: | 4.75 |
| Average bulge length [?]: | 1.75 |
| Average internal loop length [?]: | 2.3 |
| Average multi loop length [?]: | 2.6 |
| Average band length [?]: | 5.5 |

View the input to the analyser (secondary structure of this RNA molecule in bpseq format):

View the output of the analyser for this molecule:

Figure 6.

A sample view of a single RNA molecule web page.

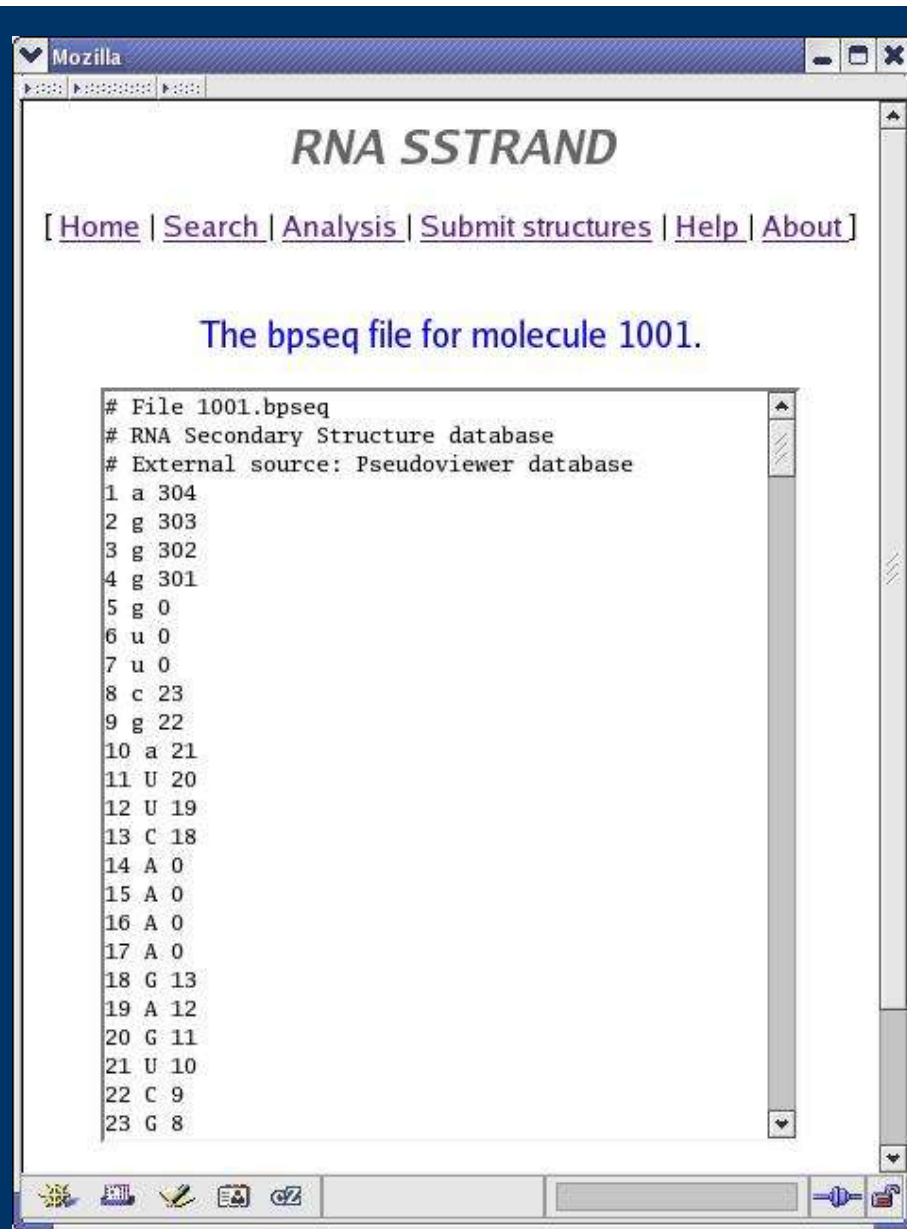


Figure 7. Input to the analyser.

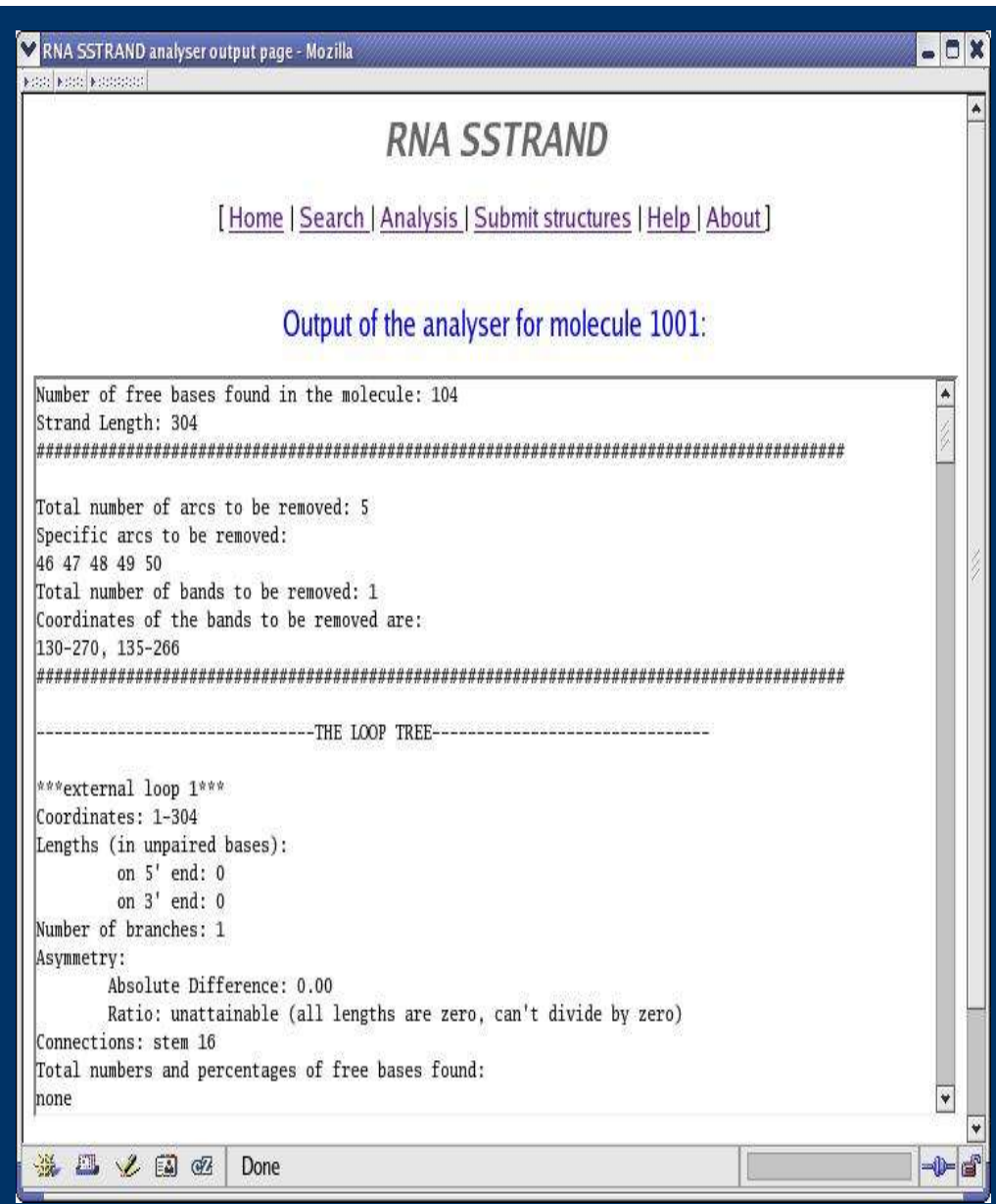


Figure 8. Output of the analyser.

Statistical Findings

- ◆ a number of interesting facts have been discovered and reconfirmed, while new directions for research have been revealed
 - ◆ previously mostly avoided, the pseudo-knotted regions have been studied more closely
 - ◆ 74.11% of structures in our database contain pseudo-knotted regions
 - ◆ plotting of a certain domain (stem, multi-loop, pseudo-knotted region, etc) versus the length of a molecule revealed a nearly perfect correlation
-
-

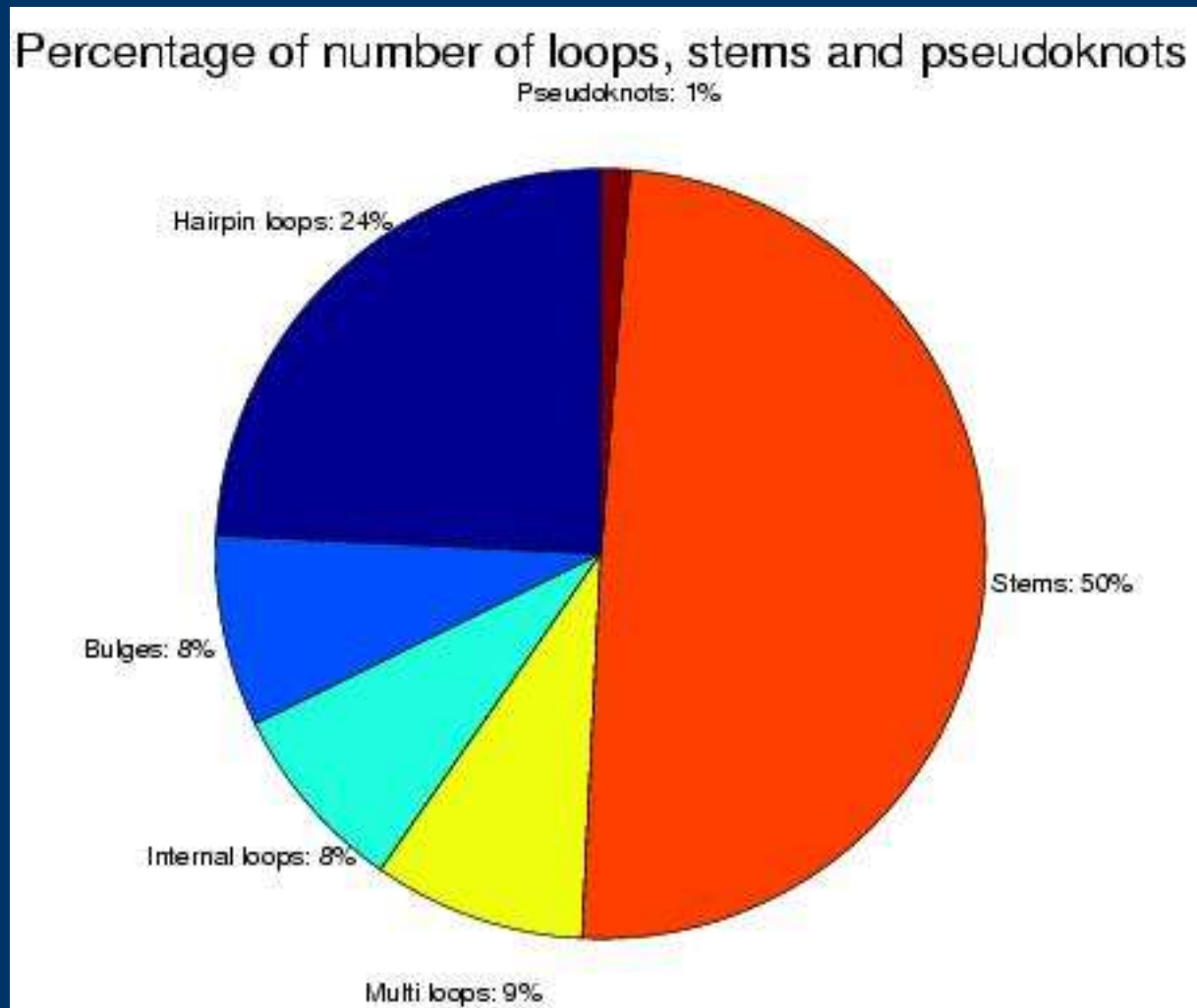


Figure 9. Fraction of each structural element in all RNA SSTRAND entries, excluding tRNAs (as they comprise the bulk of the entries and may bias the plot)

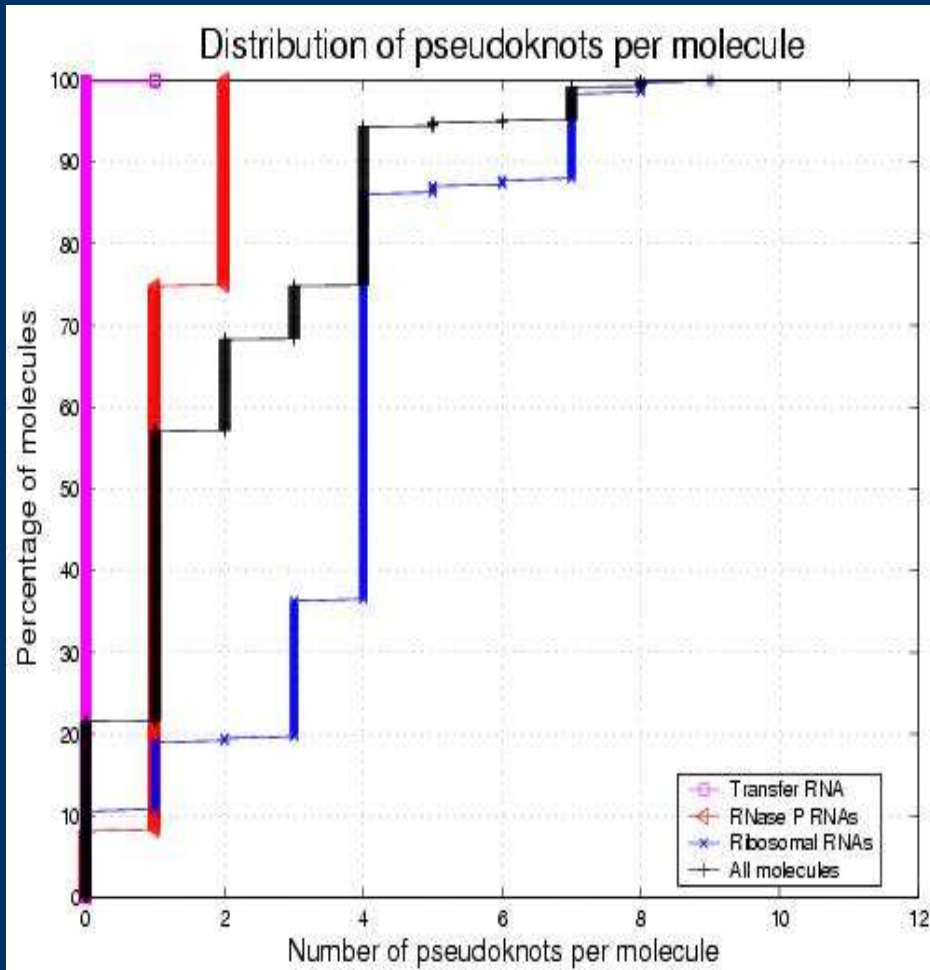
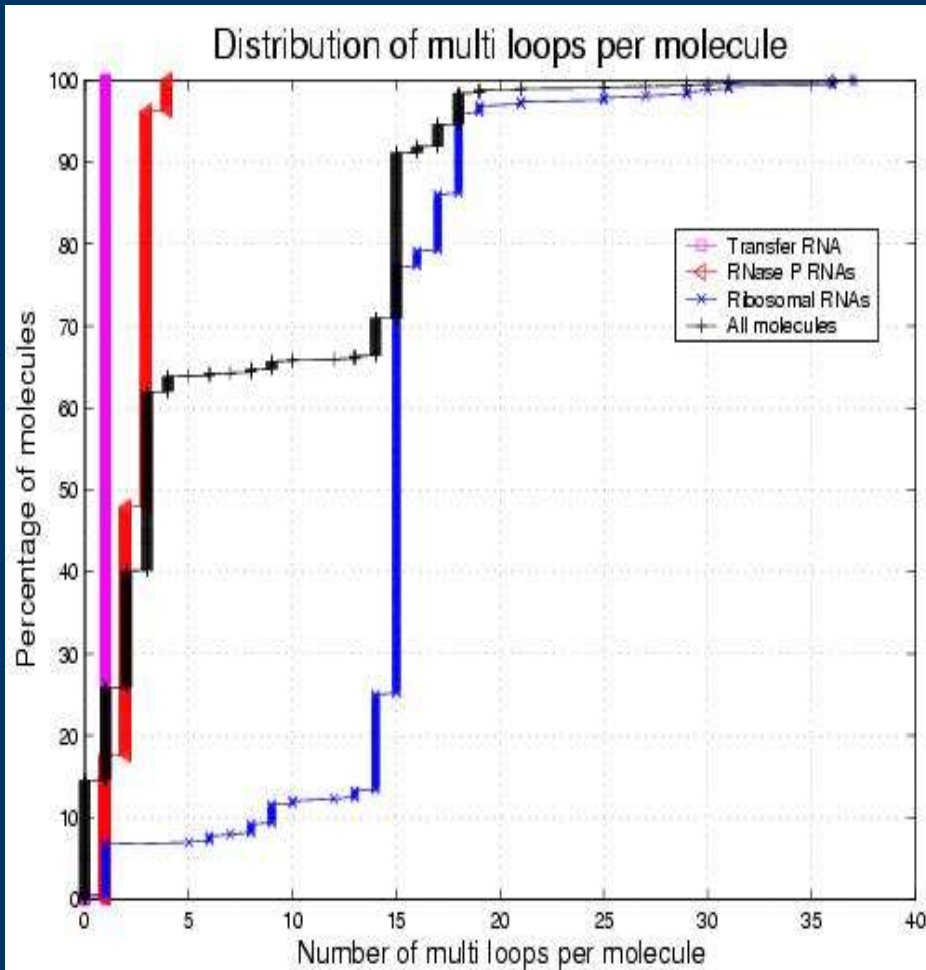


Figure 10.

Figure 11.

It might be interesting to compare pseudo-knotted versus non pseudo-knotted regions and their distributions per molecule.

Future Work

- ◆ add more data to RNA SSTRAND
 - ◆ expand the search functionality with additional structural criteria
 - ◆ support structural analysis of user-selected sets of RNAs, as obtained in the form of search results
 - ◆ use the information and functionality currently provided by the database as a starting point for studying weaknesses of current thermodynamic models and prediction methods for RNA secondary structures
 - ◆ ultimately improving these models and prediction algorithms
-
-

Acknowledgements and References

- ◆ [1] Apostolico, Alberto, Atallah, Mikhail J., Hambruch, Susanne E.(1996) New clique and independent set algorithms for circle graphs. *Discrete Applied Mathematics*, 32(1):1—24.
- ◆ [2] Condon, Anne, Davy, Beth, Rastegari, Baharak, Tarrant, Finbarr, Zhao, Shelly. (2004) Classifying RNA Pseudoknotted Structures, *Theoretical Computer Science*, 320(1):35—50.
- ◆ for more acknowledgements and references, please, refer to the paper about the RNA SSTRAND that will be published in the upcoming January database issue of the *Nucleic Acid Research* journal.