



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.
- ◆ Analyser makes use of two different algorithms:
 - Condon, Anne, Davy, Beth, Rastegari, Baharak, Tarrant, Finbarr, Zhao, Shelly. (2004) *Classifying RNA Pseudoknotted Structures*. Theoretical Computer Science, 320(1):35—50
 - 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

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 single RNA molecule/several groups of molecules
 - ◆ 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
 - ◆ single molecule page:
 - quick statistics about the molecule
 - two buttons to view input/output of the analyser for this molecule
-
-

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 1.

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					

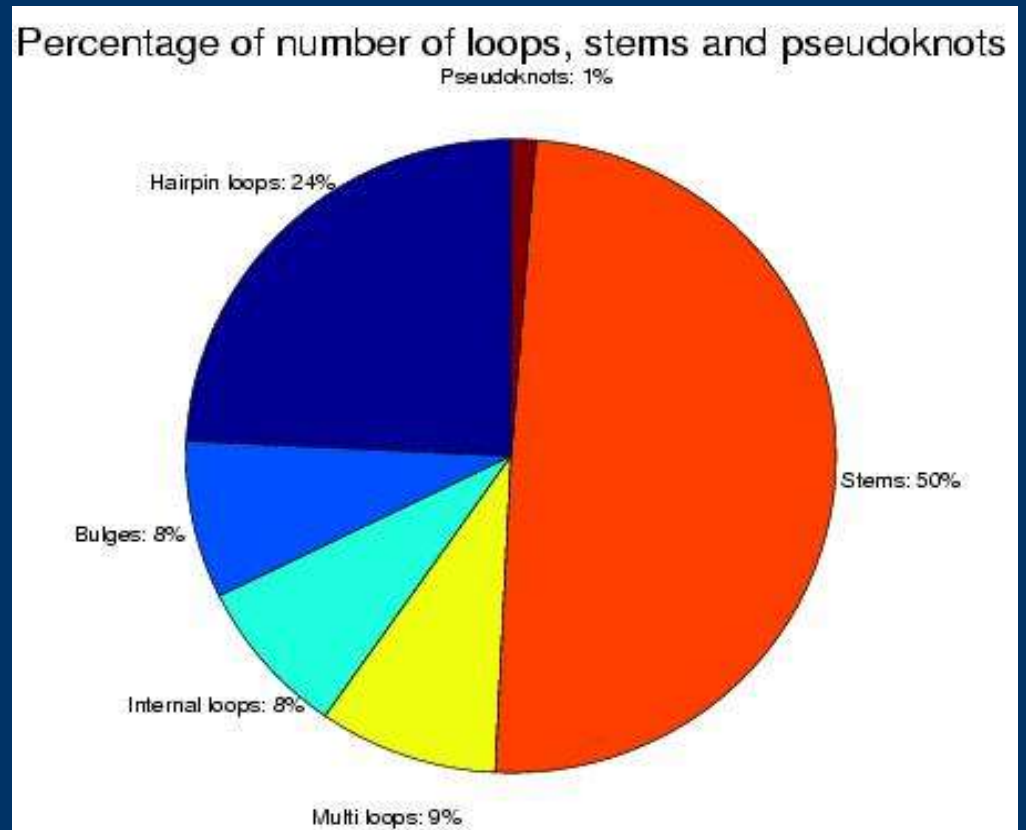
Done

Figure 2. A sample of search results.

Statistical Findings

- ◆ a number of interesting facts have been discovered and reconfirmed, while new directions for research have been revealed
- ◆ plotting of a certain domain (stem, multi-loop, pseudo-knotted region, etc) versus the length of a molecule revealed a nearly perfect correlation

Figure 3. 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)



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
-
-