

RNArobo 1.9 – Quick Start Guide

RNArobo is an RNA structural motif search tool. RNArobo can search sequence databases in FASTA format for a motif defined by a ‘descriptor’, which can specify primary and secondary structure constraints.

The format of an RNArobo descriptor is an extension of the descriptor format used by RNABob (a search tool by Eddy [1996]), so that RNABob descriptors are compatible with RNArobo. The command line options of RNArobo are also similar to those of RNABob.

Unlike RNABob, RNArobo enables you to allow nucleotide insertions in a structural element. Syntax of allowing insertions is similar to specification of the maximal number of mismatches (or mispairs). You can specify the maximal number of insertions and what nucleotides are allowed as an insertion. Insertions are not allowed at the very beginning and end of the matched regions and in a helix insertions must not be adjacent nor opposite. Usage should be clear from the example descriptor:

```
h1 s1 h1'
h1 0:0:2 NNN**CC:GG**NNN:A
s1 0:1 ACT:R
R s1 h1
```

This motif is composed of two elements – one helix and one single strand. In the helix we allow up to 2 insertions of Adenosine, while in the single strand only one insertion of a purine nucleotide is allowed (‘R’ stands for Guanine or Adenosine).

The last line of the example descriptor above illustrates usage of an optional reorder command which specifies the order in which elements are internally searched by the RNArobo algorithm, similarly to RNAMot [Gautheret et al., 1990]. If this command is absent or does not contain all elements, it is automatically supplemented by all remaining elements in order calculated by an unsophisticated heuristic which prefers longer elements with fewer wild cards. This command has no impact on the results of the search, but specifying ‘good’ order can speed up the search. The goal is to search first for the most specific elements of the motif, while leaving the loose elements to be searched at the end.

To run RNArobo on your system, you need GCC C++ compiler (tested with version 4.4.5) or for 64-bit Linux systems we directly provide an executable binary.

1. **Download** the most recent version of RNArobo at <http://compbio.fmph.uniba.sk/rnarobo>. There you can download the executable binary for 64-bit Linux systems as well as the source code package.
- 2a. If you are going to use the provided binary, set the ‘**executable bit**’ by command:

```
chmod a+x rnarobo-1.94-linux64
```

Now you are ready to run RNArobo. In what follows we refer to the binary as ‘rnarobo’, so please substitute ‘rnarobo-1.9-linux64’ for ‘rnarobo’ where applicable.

- 2b. If you cannot use the provided binary, you have to **compile RNArобо** from the source code on your own. For this step you need to have GCC compiler installed. First unpack the downloaded package, than go to the unpacked directory and execute "make" command.

```
tar -zxf rnarobo-1.94.tar.gz
cd rnarobo-1.94/
make
```

By now, you should have an executable file called 'rnarobo'.

(Optional) To install rnarobo to be available from every directory, execute "sudo make install" (you will need to enter the superuser password).

3. **Run RNArобо** by the command:

```
./rnarobo <descriptor-file> <sequence-file>
```

where '<descriptor-file>' is the path to the descriptor file and '<sequence-file>' is the path to the sequence database in FASTA format. If rnarobo is properly installed, you can run it from every directory by the same command, but without the './' prefix.

If you also want to **search in the complementary strands** of the sequences, run RNArобо with '-c' flag:

```
./rnarobo -c <descriptor-file> <sequence-file>
```

Output of an RNArобо run is printed on the standard output and consists of a header and of a list of found matches. Matches in the list are in the order as they were found in the database file from its beginning to its end. Every match is compounded of two lines. The first line gives the name and description (if any) of the sequence where this match occurs, the beginning position where the match starts in the sequence and the ending position where the match ends. This line is followed by a line containing the match itself, that is, the substring of the sequence defined by the starting and ending positions. A symbol of pipe '|' delimits individual elements of the match.

References

- S.R. Eddy. RNABob: a program to search for RNA secondary structure motifs in sequence databases. unpublished, 1996.
- D. Gautheret, F. Major, and R. Cedergren. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput Appl Biosci*, 6(4):325–31, 1990. ISSN 0266-7061.