MODENA ver. 0.0.67

A. Taneda

12 August 2014

1. Requirements

Please install RNA structure prediction methods/package by which you would like to design RNA sequences. For example, it is necessary to install the Vienna RNA package if you would like to design an RNA sequence which folds into your target structure when its structure is predicted with RNAfold [1]. I have tested the following prediction methods with the current version of MODENA: Vienna RNA 1.8.3 [1], RNAstructure 5.3 [2], CentroidFold 0.0.9 [3], UNAFold.pl 3.8 [6], IPknot 0.0.2 [4], NUPACK 3.0 [5], HotKnots 2.0 [8], pknotsRG 1.3 [7] (if the output format of each method does not changed, the other versions are also applicable as *method*). In addition to the Vienna RNA package, accessFindPath.py which can be downloaded at the RNAtabupath website [9] is also necessary if you would like to use method FindPath for predicting the energy barrier height of the designed sequence.

PLEASE CONFIRM THAT an environment variable for each structure prediction method is set appropriately, if necessary.

2. Installation

Copy an executable file of MODENA to a directory to which a PATH is set. To use RNAstructure or HotKnots/computeEnergy as a prediction method, copy modena_dps4RNAstructure.pl or modena_dps4HotKnots.pl also to the directory.

3. Usage

After the installation, type the following command:

```
modena -f example.inp
```

Designed sequences are output to the standard output. A Help for command line options is output by typing 'modena -h'. To design an RNA sequence according to your purpose, it is necessary to edit a target file (input file) for MODENA. MODENA automatically generates a working directory (and deletes the directory after the design finished, if possible).

4. Recommended usage

The easiest way to design multistable RNAs by MODENA is as follows: (step 1) select a template input file from the template directory in accordance with your purpose (e.g. a three-target design, ribozyme-based RNA device design, etc.); (step 2) edit the input file to

replace the target structures in the input file by your target structures; if you would like to design RNA device, it is necessary to specify sequence constraints and secondary structure constraints too; (step 3) execute MODENA.

5. Target file

The target file (input file) of MODENA is comprised of three components: target structures, objective functions, and a *method* list. The following is an example of two-target design without sequence constraints:

```
.((((((((((.((((((.........)).)))).....)))))))...)))).......
..((....))...((((((.........)).)))))..((((((((((....))))))))))).

;
-1*((F:CONT-50)^2)^0.5
-1*(C:FE-B:EFE+D:FE-B:EFE)
-1*((D:FE-C:FE-0.0)^2)^0.5
;
B RNAfold-p 1 "-d2"
C RNAeval 1 "-d2"
D RNAeval 2 "-d2"
F GC 0
```

The first two lines are target structures in the bracket notation. The next blank line indicates no sequence constraint is used for this design. A ';' is a separator.

5.1 Objective function

The next three lines are expressions specifying the objective functions, where one line corresponds to one objective function. A letter just before a ':' indicates *method* and the keyword just after a ':' specifies the *property*, e.g. FE means a free energy of a target structure and EFE means an ensemble free energy. IT IS NOTED THAT MODENA EXPLORES SEQUENCES MAXIMIZE THESE OBJECTIVE FUNCTIONS. If you would like to minimize some objective functions, a -1 must be multiplied to the objective function in the expression. In these expressions, we can use *properties*, numerical values, and +-*/^(). IT IS NOTED THAT the number of objective functions must be larger than or equal to two (i.e. a single objective function is not allowed). If you would like to perform such a design, please use modena.pl (see 7.1 modena.pl).

5.2 *Method* list

The last component is a *method* list. The available *methods* and *properties* are summarized in Table 1 of the paper of MODENA for multistable RNA design. These lines specify secondary structure prediction methods for evaluating each individual during the GA optimization. In each line, the leftmost letter is a variable assigned to the *method*. As the variable, you can utilize an alphabetical letter from A to Y and from a to z (a capital Z is not allowed). For example, if you would like to use the value of property FE predicted by method A in an objective function, you can use A:FE in your expression of the objective function.

Since positive and negative operators use a secondary structure predicted by a *method*, at least one *method* which can predict a structure (the methods whose 'str.' column in Table 1 is marked by 'y') should be included in the *method* list. If you wouldn't like to include such a method in your *method* list, MODENA options –opPos 0 –opNeg 0 should be used not to invoke positive and negative operators.

5.2.1 Specification of target structure

Some prediction *methods* need a number indicating a target structure. For example, the *method* at the third line in the *method* list,

```
C RNAeval 1 "-d2"
```

computes the free energy of target structure 1. All *methods* except the following three *methods* need the specification of a target structure: pfunc, EnsembleEnergy, FindPath. *Method* FindPath needs two secondary structures. E.g., the barrier between target structure 1 and target structure 2 is computed by

```
FindPath 1 2 1000
```

Method FindPath can utilize not only numbers specifying target structures, but also *method* variables to specify the predicted structures. The barrier between two secondary structures predicted by *method* B and *method* C is computed by

```
FindPath B C 1000
```

where a look-ahead parameter = 1000 is used [9]. It is noted that user can specifies the only *method* variables which appear above the *method* in the *method* list.

5.2.2 How to specify options of each prediction method

All *methods* except for FindPath, command line options can be invoked by specifying the options in the " " just after the number indicating a target structure; e.g. in the above example target file, -d2 option is specified for method B, C, and D. IF THE OUTPUT FORMAT OF THE *METHOD* DOES NOT CHANGE, we can freely specify the options of each method

through the " ". If you do not need any option for the *method*, the " " can be omitted.

5.2.3 Secondary structure constraints

To use secondary structure constraints for a *method* (Fold, CentroidFold or RNAfold in the current version of MODENA), add an & to the right side of the *method* line; then you can specify secondary structure constraints in the next line. An example is shown below:

```
C Fold 2 &
((((....))))**********************
```

The positions specified by * or ? are not constrained. IT IS NOTED THAT the structure constraints must be consistent with the target structures in terms of base complementarity (the dependency graph) and sequence length.

5.2.4 Sequence masking

If you would like to use a sequence masking for a *method*, a + is added to the right side of the *method* line; your masking sequence can be specified in the next line. E.g., by using the following two lines in the method list, we can fold the designed sequence after replacing the right five nucleotides by NNNNN:

```
A RNAfold 1 "-d2" +
.............................NNNNN
```

The positions specified by . are not masked. The length of the masking sequence must be the same with that of the target structures.

6. Output format

An example of output is shown below:

```
Individual= 11  Rk= 1  Sc= -3.571430 -8.880001 -7.200001
ACAGUUCGAUCCCGAACGCUAAUACGAAGUAUCAGCGUUACUGAUGCACUUCGUUG
B RNAfold-p MFE:-17.100000 SIM:1.000000 EFE:-17.940001 PB:0.255320
C RNAeval FE:-17.100000
D RNAeval FE:-9.900000
F GC CONT:46.428570 PAIR:0.000000
...(((((....))))).....((((((((((.....))))).))))))).. comm:B
Tract>3nt= 0 Constr= 0 tractType= ACGU
```

The first line contains an individual number, the dominance rank, and the values of objective functions for the designed sequence. The individual number can range from 0 to 2×(population

size)-1. The second line is the designed sequence. The lines having *method* names below the designed sequence line show the *property* values predicted by each *method*. The line with the "comm:B" on the right side indicates the secondary structure predicted by *method* B. The last line gives information about detected nucleotide tracts. In this example, no tract has been found. If a tract exists and the tract overlaps sequence constraints, "constr=" indicates how many nucleotides are overlapped. If undesired sequence motif option is invoked, information about detected undesired sequence motif is also output here.

7. Utility scripts

7.1 modena.pl

By using modena.pl, the user can specify maximization/minimization of each objective function. To specify max./min., a "max" or "min" keyword is added to the right side of the expression of an objective function (a space character between the expression and the keyword is mandatory). The following is an example input file for modena.pl (this example is equivalent to the example input shown in the "5. Target file" section):

```
.((((((((((.((((((.........)).)))).....)))))))...)))).......
..((....))...((((((.........)).)))))..((((((((((....))))))))))).

;
((F:CONT-50)^2)^0.5  min
C:FE-B:EFE+D:FE-B:EFE min
((D:FE-C:FE-0.0)^2)^0.5 min
;
B RNAfold-p 1 "-d2"
C RNAeval 1 "-d2"
D RNAeval 2 "-d2"
F GC 0
```

Moreover, modena.pl can perform the design with a single-objective function. It is noted that NSGA2 utilized in MODENA may be a not so good choice for the optimization with a single objective function.

7.2 modena_nonred.pl

MODENA can output identical RNA sequences for a design (this is the case when a too large GA population size is specified). After saving the output into a text file (e.g. example.out), we

can check whether identical sequences are included in the output file or not through the following command:

```
./modena_nonred.pl example.out
```

The utility script (modena_nonred.pl) is contained in the utility directory.

Reference

If you use MODENA, please cite the following paper:

A. Taneda, Multi-objective optimization for RNA design with multiple target secondary structures, submitted.

OF COURSE, PLEASE CITE THE STRUCTURE PREDICTION METHOD USED FOR YOUR DESIGN, TOO!

[1] Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie Chem. Mon., 125, 167–188.

[2] Reuter, J. S. and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics, 11, 129.

[3] Hamada, M., Kiryu, H., Sato, K., Mituyama, T., and Asai, K. (2009) Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics, 25, 465–473.

[4] Sato, K., Kato, Y., Hamada, M., Akutsu, T., and Asai, K. (2011) IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. Bioinformatics, 27, i85–i93.

[5] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011) NUPACK: Analysis and design of nucleic acid systems. J. Comput. Chem., 32, 170–173.

[6] Markham, N. R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. Methods Mol. Biol., 453, 3–31.

[7] Reeder, J. and Giegerich, R. (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC Bioinformatics, 5, 104.

[8] Andronescu, M., Aguirre-Hern´andez, R., Condon, A., and Hoos, H. H., (2003) RNAsoft: A suite of RNA secondary structure prediction and design software tools. Nucleic Acids Res., 31, 3416–3422.

[9] Dotu, I., Lorenz, W. a., Van Hentenryck, P., and Clote, P. (2010) Computing folding pathways between RNA secondary structures. Nucleic Acids Res., 38, 1711–1722.