Folding RNA/DNA hybrid duplexes

Ronny Lorenz1,*, Ivo L. Hofacker1,2 and Stephan H. Bernhart2,*

1Institute for Theoretical Chemistry, University of Vienna, 2Research group BCB, Faculty of Computer Science, University of Vienna, Währinger Straße 17, 1090 Vienna, Austria and 3Department of Bioinformatics, University of Leipzig, Haertelstraße 16-18, 04109 Leipzig, Germany

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: While there are numerous programs that can predict RNA or DNA secondary structures, a program that predicts RNA/DNA hetero-dimers is still missing. The lack of easy to use tools for predicting their structure may be in part responsible for the small number of reports of biologically relevant RNA/DNA hetero-dimers.

Results: We present here an extension to the widely used ViennaRNA Package (Lorenz et al., 2011) for the prediction of the structure of RNA/DNA hetero-dimers.

Availability: http://www.tbi.univie.ac.at/~ronny/RNA/vrna2.html

Contact: ronny@tbi.univie.ac.at, berni@bioinf.uni-leipzig.de

Supplementary Information: Supplementary data are available at Bioinformatics online.

Received on April 10, 2012; revised and accepted on July 19, 2012

2 APPROACH

2.1 Adaptation of the ViennaRNA package

For the prediction of RNA/DNA hetero-dimers, three distinct energy parameter sets (RNA, DNA and mixed) are necessary. In order to keep the changes within the inner recursions of the algorithm minimal, we use 8 instead of 4 bases and 24 instead of 6 types of Watson–Crick or wobble base pairs for the energy computations. The look-up tables for the energy parameters grow accordingly. However, this does not significantly contribute to memory consumption. Finally, we demand a fixed order for the input sequences: RNA first, DNA last.

2.2 Energy parameters

Since we are only considering RNA/DNA hetero-dimers, not co-polymers, we only need a subset of the energy parameters necessary for the full computations in the RNA or DNA case: we do not encounter stacking of a RNA on a DNA base, for example. In contrast, stacking parameters of RNA/DNA base pairs, multi-loop penalties and the interior loop parameters are needed. For stacked pairs, i.e. perfect helices, the parameters have been determined experimentally (Martin and Tinoco, 1980; Sugimoto et al., 1995). Stacking parameters are the most important part of the energy model, as they account for the majority of RNA/DNA binding free energy. Unfortunately, for other mixed RNA–DNA loops no experimental data are available.

As a final biological example, DNA/RNA hybrids play a role within the CRISPR pathway in bacteria and archaea (Howard et al., 2011). On a technical side, RNA/DNA dimers are often used in micro-array or PCR experiments. The ability to computationally investigate the properties of the structures of interacting RNA and DNA molecules would help in the analysis of these dimers. While a number of programs exist that can predict RNA or DNA secondary structures (For a review, see e.g. Reeder et al., 2006) or even their dimers, there is, to our knowledge, no program that also includes the possibility to predict the full secondary structure of RNA/DNA hetero-dimers. This may be due to the lack of compiled energy parameters, as there are only stacking energies available today. We introduce here a possibility to predict the secondary structure of such hetero-dimers within the widely used ViennaRNA Package.

*To whom correspondence should be addressed.
parameters (Supplementary Fig. S1), so we simply use the average of RNA and DNA $E_{RD} = (E_R + E_D)/2$ for all missing values. We are aware that this formula gives an at best crude approximation of the actual folding energies, but note that new parameters can easily be incorporated. We hope that the availability of prediction tools for RNA/DNA hetero-dimers will encourage experimentalists to provide the missing energy parameters.

2.3 Inclusion into the ViennaRNA package

Our approach to just adapt the energy computations enabled us to easily include the RNA/DNA hetero-dimer support into all parts of the ViennaRNA Package that compute dimer structures, namely, RNAcofold (Bernhart et al., 2006), RNAup (Mückstein et al., 2008) and RNAduplex (Tafer and Hofacker, 2008). RNA/DNA hetero-dimers are supported for both the minimum free energy as well as the partition function-based calculations. Furthermore, all programs can read in user-supplied energy parameters. In particular, this allows to make immediate use of any newly determined parameters for mixed RNA–DNA loops, or enables scientists to easily evaluate different energy models as soon as sufficient RNA/DNA structure data are available.

3 RESULTS

In the absence of programs that can deal with RNA/DNA dimers, researchers have resorted to simply treating both parts of the dimer as RNA. To test the effect of applying the right (DNA or RNA) parameters to the single molecules and using explicit RNA/DNA interactions, we chose ANRIL, a long non-coding RNA that is suspected to have a DNA target sequence, and one of its putative targets. In order to simulate the size of the transcription bubble, we chose a sliding window approach with DNA pieces of length 50, step size 25 for our computations. Even though ANRIL (1500 nt) is much longer than the pieces of DNA bound to it, we sometimes observe a dramatic effect on the structure of the hybrid molecule. The tree edit distance of the hybrid structures predicted with RNA parameters and with our hybrid approach rises up to 900, and the structure of the binding site can vary greatly (Fig 1).

Another example for the usefulness of the mixed parameters is the strand dependency of R-loop formation (Reddy et al., 2011). A rGAA trinucleotide repeat can form R-loops while its reverse complement r(UUC)$_n$ cannot. The non-symmetric mixed parameters can explain that: addition of another trinucleotide will give $-4.1$ kcal/mol for rGAA, but only $-2.6$ kcal/mol for r(UUC). In contrast to RNA or DNA parameters, the mixed parameters can also explain the changes in R-loop formation of other trinucleotides investigated (see Supplementary Fig. S2).

Funding: This work was funded by the Austrian 'SFB 43 RNA regulation of the transcriptome'. This work was funded by the Austrian GEN-AU project ‘Bioinformatics Integration Network III’.

Conflict on Interest: none declared.

REFERENCES


