Tutorial - 2D/3D RNA Structure

From 1D to 2D & 3D Structures of RNA

Abstract

This tutorial will introduce several RNA Visualization and computational approaches that allow the prediction of RNA structure from sequence. We will briefly overview the main sources of RNA data, and introduce the main data formats used to encode structural information. Tutees will have the opportunity to gain practical experience in using RNA structure prediction software and visually evaluate their output against known RNA 3D structure .
Tools that will be covered:

- Structure prediction: RNAFold, RNAalifold, LocARNA.
- Visualization and assessment: VARNA, Pseudoviewer, VIENNA web server, Jalview
- 3D Structure annotation: RNAView, Jmol
- Required software: Java + any Web browser

Tentative schedule

9.30 - Intro
9.45 - First session:

- databases
- 3D structure Annotation
- hands on
- intro structure prediction
- begin hands on

10:30 - Coffee
11:00 - Second session

- hands on structure prediction
- assessment

12:30 - discussion and other topics.
13:00 - finish

Detailed program (Slides)

Intro

Representations of RNA: 1D, 2D, 3D
Getting 3D out of the picture as quickly (and diplomatically :)) as possible

First contact with file formats and visualizations

Retrieve data from non-coding RNA Databases:

- sequences
- structures
- alignments
- experimental data

Quickly visualize them using VARNA.

Minimal Free-Energy folding of RNA

- Fetch the D1-D4 domain of the Group II intron (RFAM ID: RF02001, Seed alignment)
- Load it into VARNA
- Extract the *A. Capsulatum* (id:Acidobacterium_capsu.1) sequence.

A. Capsulatum sequence of a group II intron RNA

```
ACCGUCGCGAGGCGGGGUUUGAAGGAAGCGUGGAGCAAAUGCGCGAGUCGAUGGACAAAAACCGGAUACAAGGCGU
GAGCGUAGGGCGAGCGGGCAACUUACCGCGAAGCCCAUAUCCAUUAAGUACGCGGAGCGUAGAUCCGGCGACUGUG
CGCAGAAGGCGAUUGAGC
```

- Run RNAFold on this sequence using the Vienna RNA web tools suite
- Visualize the prediction within VARNA and compare it with the Native secondary structure using use the algorithms layouts (Ctrl+1-4).
- Rerun RNAFold using more recent energy parameters (*Show advanced options* > *Turner 2004 energy model*)
- Compare the quality of the prediction.

Automated schematic annotation of full-atom 3D models

- Retrieve a complete, *all-atoms*, structural model of the Group II intron (PDB ID 3IGI)
- Annotate PDB model using RNAview (Option -p to produce an RNAML annotated file)
- Visualize the structure within VARNA
- Run RNAFold on the 3IGI sequence, and compare its prediction to the base-pairs inferred by RNAView from the 3D model.

Pseudoknots

The modest results obtained in our previous experiment are due to the absence of crossing interactions, or *pseudoknots* in the computational scheme underlying RNAFold. Although the prediction of such features is computationally expensive, simplified approaches have been recently developed:

- Fetch sequence/structure data for a domain of the tmRNA from the PseudoBase++ (PseudoBase ID: PKB210)

Pseudoknot PKB210 of a transfer-messenger RNA (tmRNA)

```
CCGCUGCACUGAUCUGUCCUUGGGUCAGGCGGGGGAAGGCAACUUCCCAGGGGGCAACCCCGAACCGCAGCAGCG
ACAUUCACAAGGAAU.[[[[[[..[[[...{{{{{{{..]]].((((((((((....))))))).((((....)))
)...)))).]]]]]].......}}}}}}}.
```

- Visualize it using the Pseudoviewer web server or VARNA.
- Predict its MFE using RNAFold, and visualize the result
- Use PKnots-RG software to predict its structure (blind *mfe* mode, and enforcing the

presence of a pseudoknots), and visualize the prediction

Visualize ensemble features

Sometimes, considering the average conformation in the Boltzmann weighted ensemble yields a better prediction than the minimal-free energy structure returned by RNAFold. Even in situations where a centroid-base method, based on this principle, does not return a good candidate, the Boltzmann probability can still be interpreted as a confidence score for the prediction method.

S. cerevisiae Phe-tRNA
>ENA|M10740|M10740.1 Saccharomyces cerevisiae Phe-tRNA. : Location:1..76 GCGGATTTAGCTCAGTTGGGAGAGCGCCAGACTGAAGATTTGGAGGTCCTGTGTTCGATCCACAGAATTCGCACCA

Indeed, folding the tRNA sequence above gives a very poor prediction, completely missing one of the helices. However, looking at the base-pairing probability matrix (BP dot-plot), one observes the presence of the overlooked helix within the suboptimal structures. In general, BP dot-plots can be used to assess one's confidence in predicted structures.

Adding comparative data improves prediction

- The sequences below were prepared by:
    - Retrieving the seed alignment (ungapped FASTA format) in RFAM for the SAM riboswitch (RFAM ID: RF00162) from Jalview (launch development version).
    - Copying first 6 sequences to new alignment and remove gaps.
    - Adding the sequence from PDB model 2GIS (file->fetch sequences .. from PDB)
- Feed the 7 sequences below to LocARNA.
- While the web server runs, visualize the consensus, and the RNAView annotated 3D model structure within VARNA.
- Use VARNA to :
    - Compare the raw predictions of locARNA with the RFAM consensus structure.
    - Do the same for the consensus prediction output by RNAAlifold.
- Open the locarna result in a text editor, and replace the ***lower*** secondary structure line in the alignment with the '*alifold'* consensus line.
- save the file and load it into Jalview (launch development version)
    - compare the different structure predictions.
        - on the alignment (single and base pair consensus)
        - in an embedded VARNA window (right click on a sequence, follow 'structure->view structure submenu)
    - view the 2GIS structure and colour it by RNA helices
        - follow structure menu to add PDB ID 2GIS, and then open it via the 'View Structure' menu

First 6 sequences of the RFAM RF00162 seed alignment + PDB 2GIS

```
>Bacillus_liche_1/1-110
UCCUUAUCAAGAGUGGUGGAGGGACUGGCCCUGUGAAACCCGGCAACCGCUGUCUAUGACAGAAUGGUGCUA
AAUCCUUAAGAGCAUGUUCGUGCUCUUGAAGAUAAGGA
>Bacillus_licheni/1-100
UUCUUAUCAAGAGCAGGCAGAGGGACAAGCCCGAUGAAGCCCGGCAACCGACUUUUUAAGCACGGUGCUAAU
UCUUGCAGCUGACGCUGAGAGAUAAGGA
>L_innocua_6_1_11/1-119
AUCUUAUCCAGAGUGGUGGAGGGAAAUGCCCUGUGAAACCCAGCAACCUAAACAAUAAUUCAUUAUGUGUUU
AAGGUGCUAAGUCAUGCAGAACAACGAUUUGUUCUGAAAGAUGAGAA
>L_innocua_7_1_10/1-109
CUCUUAUCCAGAGCGGUAGAGGGACUGACCCUUUGAAGCCCAGCAACCUACACAUAUAAGUGAAAGGUGCUA
AUCUGUUGCAGGAGUAAUAUCUCCUGAACGAUGAGAG
>2GISA/1-94
GGCUUAUCAAGAGAGGUGGAGGGACUGGCCCGAUGAAACCCGGCAACCAGAAAUGGUGCCAAUUCCUGCAGC
GGAAACGUUGAAAGAUGAGCCA
>Pelobacter_propi/1-107
UGCUUAUCAAGAGUGGUGGAGGGAAAGGCCCUGUGAAACCACAGCAACCGGCCGUCAUGGCGCCAGGUGCUA
AUUCCUGCCGUCAGGCAAAGAUGAGAGGGUGUGCU
>Bacillus_liche_2/1-138
CUCUUAUCCCGAGCUGGUGGAGGGACAGGCCCAAUGAAACCCAGCAACCGGUUUCUCUUAUUAAUGGAAAAA
AACAGUUUCUGAGACAACUACGGUGCUAACCUGAUGCAAGGUGUUCAAUACCUUGAGCGAUAAGAG
```

Other things to do with Jalview:
Drag the following URL onto the Desktop window:
http://www.compbio.dundee.ac.uk/user/ws-dev1/jalview/develop/applet/RF00031_folded.stk

Summary and discussion