Helices of RNAs

Jiabin Huang

AG Experimentelle Bioinformatik (Cyanolab)
Institut für Biologie III
Universität Freiburg

February 22, 2011

Outline

- Introduction
- Development of a new structure abstraction
- Implementation of an algorithm based on the new abstraction
- Possible problems
- Evaluation of the algorithm
- Designing RNA class predictors

Structural components of RNA

RNA has different structural components:

- single-stranded regions (SS)
- hairpin loops (HL)
- stacking regions (SR)
- bulges on 5´side (BL) or 3´side (BR)
- internal loops (IL)
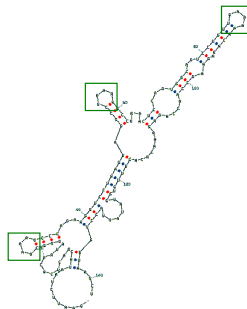- multiloops (ML).

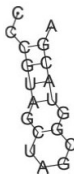# Structural components of RNA
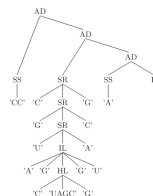

Figure: Structural components

Visualisation of structures

CCCGUAGCUAGCGGUACGA

squiggle plot

tree

Suboptimal structures

- But the ´84true" structure is not always the one with
- the lowest predicted free energy.

Introducing abstract shapes

- Solution: Use abstract shapes to describe a set of structures.

Introducing abstract shapes

- In the domain of shapes, we care only about
- open structures (OP)
- closed structure (CP)
- branching ("fork", FK) and
- adjacency of structures (AD).

Defining abstract shapes

- A (very abstract) abstraction function :
- $\pi(SS(l)) = OP$ (single-stacked region)
- $\pi(HL(a,l,b) = CL$ (hairpin)
- $\pi(SR(a,x,b)) = \pi(x)$ (stacked region)
- $\pi(IL(a,l,x,l\acute{b}4,b)) = \acute{f}0(x)$ (interior loop)
- $\pi(ML(a,c,b)) = FK(\pi(x))$ (multiloop)
- E represents the "empty structure".
- where a, b = nucleotides, l = loop, c = list of adjacent
- components and x = arbitrary structure elements.

Example

Computing abstract shapes

- Abstract shapes are a homomorphic image of the folding
- space of a RNA sequence (same as mfe or base pair
- maximization).
- They can therefore be computed using Dynamic
- Programming (DP).
- Giegerich et al. use an extension of DP called
- Algebraic Dynamic Programming (ADP).
- ADP offers some interesting aspects (separation of
- recognition and evaluation, pair algebras).

Computing abstract shapes

- *Algebraic Dynamic Programming* defines the set of all possible solutions (e.g. foldings) using a context-free grammar.

First summary

- One abstract shape represents a family of
- similar RNA structures.
- Shapes are defined by an abstraction function
- that maps from structure to shape space.
- Shapes can be used to represent a large number of (suboptimal) foldings to obtain an holistic view of the folding space.

Applications of abstract shapes

- Suboptimal Folding
- Out of 99 tRNA sequences in Rfam only 30 had the typical cloverleaf structure as predicted mfe folding.
- Example: tRNA of *Natronobacterium pharaonis:*
- mfe structure is a hairpin with internal loops, cloverleaf structure occurs at position 104 of 199
- suboptimal structures.
- All these suboptimal folding can be represented by three abstract shapes.

Applications of abstract shapes

- Suboptimal Folding
- (demo)

Applications of abstract shapes

Suboptimal folding

Applications of abstract shapes

- Better structure prediction (than mfe folding) can
- be obtained using comparative approaches:

Applications of abstract shapes

- A possible resort :

Applications of abstract shapes

- Consensus structures with abstract shapes

Applications of abstract shapes

- Possible scoring functions:

Going comparative pays off

Comparison with Sankoff

Summary

- Abstract shapes represent disjoint classes of RNA foldings.
- Shapes are computed using (Algebraic) Dynamic Programming.
- Inspecting the abstract shape space of a sequence
- can give a quick overview of the folding space.
- Consensus folding with abstract shapes performs well.
- Choice of best abstraction function and energy range
- is important but difficult.

Conclusions and Outlook

- Other approaches to suboptimal folding exist such as statistical sampling of the folding space.
- Text representations of shapes could be used as index in structure databases to classify non-coding RNA.
- Extensions to the shape formalisms are under work
- (e.g. computation of shape probabilities, de novo prediction of non-coding RNA genes)

End

- Thanks a lot for your attention !
- Questions ?