

Helices of RNAs

Jiabin Huang

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

March 1, 2011

Outline

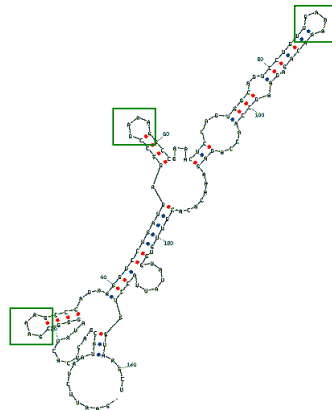
- Introducing abstract shapes
- Development of a new structure abstraction
- Implementation of an algorithm based on the new abstraction
- Possible problems

Structural components of RNA

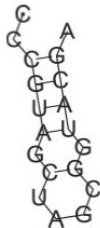
RNA has different structural components:

- 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



CCCGUAGCUAGCGGUACGA



```

graph TD
    AD1[AD] --- SS1[SS]
    AD1 --- AD2[AD]
    SS1 --- CC["'CC'"]
    AD2 --- SR1[SR]
    AD2 --- AD3[AD]
    SR1 --- C["'C'"]
    SR1 --- SR2[SR]
    AD3 --- SS2[SS]
    AD3 --- HL[HL]
    SS2 --- A["'A'"]
    SR2 --- G1["'G'"]
    SR2 --- SR3[SR]
    SR3 --- C2["'C'"]
    SR3 --- SR4[SR]
    SR4 --- U["'U'"]
    SR4 --- IL[IL]
    IL --- A2["'A'"]
    IL --- G2["'G'"]
    IL --- HL2[HL]
    HL2 --- C3["'C'"]
    HL2 --- UAGC["'UAGC'"]
    HL2 --- G3["'G'"]
  
```

tree

Suboptimal structures

But the "true" structure is not always the one with the lowest predicted free energy. So what to do?

- Enumerate suboptimal structures within a given energy range R .
- Hope to find a structure fulfilling your expectation or coming close to experimental results.

But the number of suboptimal structures grows exponentially with the energy range considered.

Introducing abstract shapes

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

- An abstract shape represents a class of similar structures sharing a common pattern of helix nesting and adjacency.
- "Abstract" since we do not care about all details of the structures.
- Each shape class has a representative structure called shrep (with minimum folding energy).

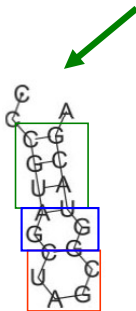
Introducing abstract shapes

In the domain of shapes, we care only about

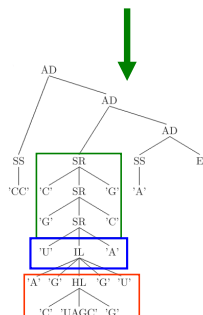
- open structures (OP)
- closed structure (CP)
- branching ("fork", FK)
- adjacency of structures (AD)

Level of detail is defined by an abstraction function π . The abstraction function can be defined on the level of the structural components.

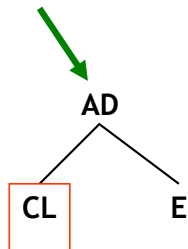
Example



squiggle plot



tree



abstract shape

Abstract shape, Energy range: 5 kcal/mol

UCGCGCACAGGACAUCCUAGGUACAAGGCCGCCGU

-6.30	.((.(..(((.....))).(((.....)))))).	[[] []]
-4.90(((.....))).(((.....))).....	[[]]
-3.90((.(..(((.....)))...)).....	[]

Drawback of abstract shape

- The major drawback of abstract shape analysis is the position independence of the abstraction

new algorithm should be:

- develop a new algorithm for RNA secondary structures which keeps track of the position of structural elements
- be implemented in the framework of "Abstract shapes of RNA"
- be evaluated and compared to existing algorithms
- The new algorithm will be used to design prediction strategies for various classes of RNAs

develop a new structure abstraction

The straightforward idea to overcome the position independence of the current available shape abstractions is

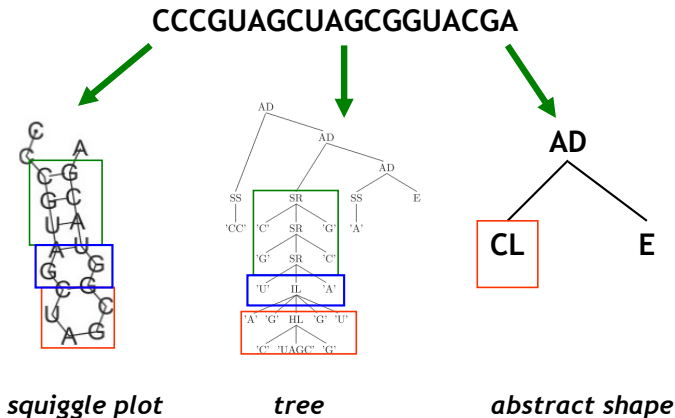
what keeps track of positions

- hairpin loop, because it is the least expensive one
- multiloops, as possible braching points, are structurally important
- stacked pairs, bulge and internal loops are the main structural contributors
- possible different helices type

which positions of the base pair will be tracked

- i
- j
- (i,j)
- $i+j/2$

Todo: capture of the pdf picture



Output from the first trial version of helices shape

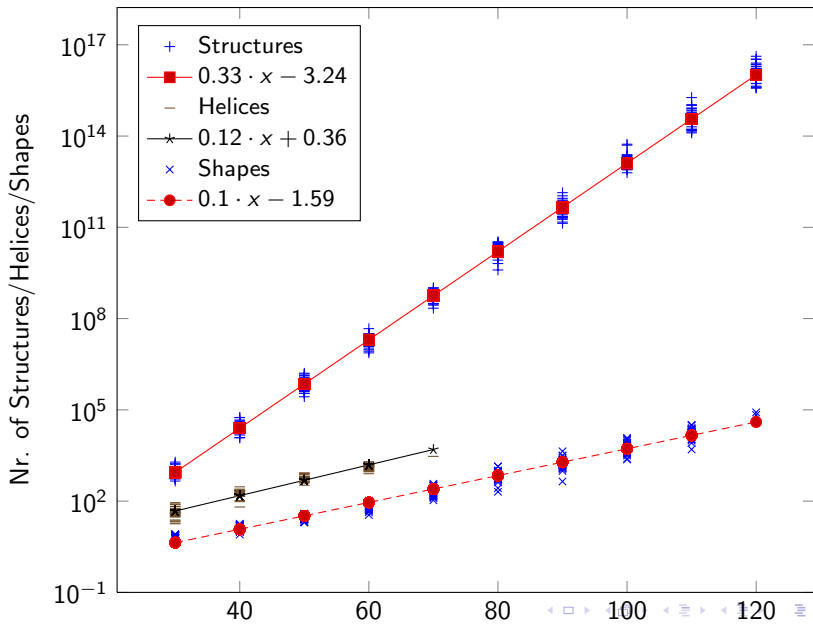
Helices shape, Energy range: 5 kcal/mol

	UCGCGCACAGGACAUCCUAGGUACAAGGCCGCCGU	
-6.30	.(((.(.(((....))).(((.....)))))).	[13.5,25]
-4.60(((....))).(((.....)))).	[13.5,26.5]
-3.90((.(((....))).)).....	[13.5]
-3.60((....(((.....)))..))..	[22]
-3.40((.(((....))).))....((....))..	[13.5,30]
-3.20	..((((.....((.....)).....)).))..	[17]
-2.80((.....(((.....))).))..	[25]
-2.40(((.....))..)	[26.5]
-1.60((.(((....))).))....((....))	[13.5,31.5]

Possible problems

- do not preserve nesting of elements and might lead to abstract shapes where a helix index occurs more than once → solved with different representation form
- abstracting from bulge and internal loops might be too rigorous → refine the definition of the abstraction, such that all critical criteria are met
- records of helices is a lot more than the records of shapes → runtime problem

Growth of structure, helix and shape space



Outlook: Designing RNA class predictors

- specific classes of RNAs, such as riboswitches or miRNA precursors, show characteristic features within their folding space
- the energy landscape of riboswitches harbors two equally low, but well separated minima, but well separated minima
- miRNA precursors is governed by one deep and well-defined minimum
- develop specialized prediction algorithms for classes of RNAs

End

- Thanks a lot for your attention !
- Questions ?