Helices of RNAs

Jiabin Huang

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

March 7, 2011

Outline

- RNA secondary structure and algorithms
- Introducing abstract shapes
- Development of a new structure abstraction
- Outlook

Structural components of RNA

RNA has different structural components:

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

Structural components of RNA
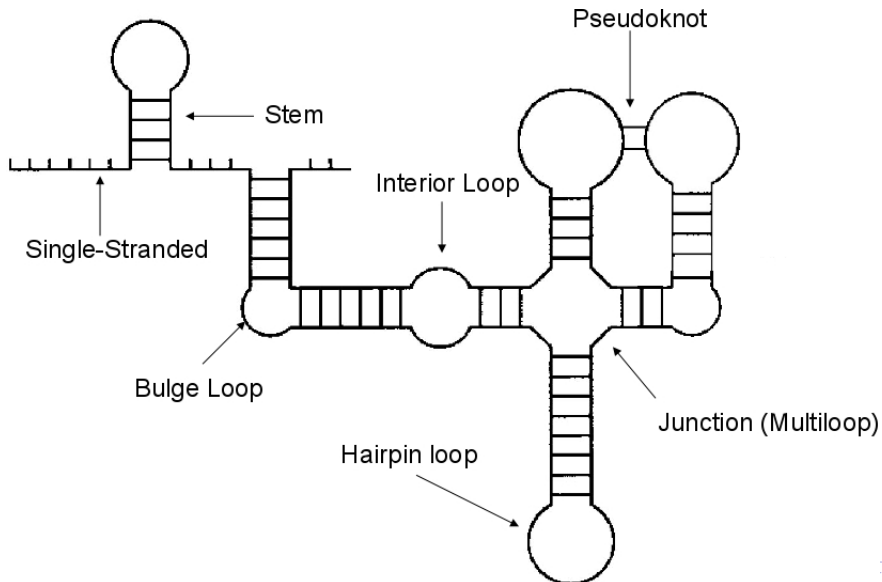
2 principles

Simplifying Assumptions for Structure Prediction

- RNA folds into one minimum free-energy structure.
- There are no knots (base pairs never cross).
- The energy of a particular base pair in a double stranded regions is sequence independent

RNA has different structural components

- single-stranded regions (first described by Zuker and Stiegler in 1981)
- RNA folding are stabilized by the stacking or by dangling end
- destabilized by hairpin, internal and bulge loops
- the runtime is $O(n^3)$

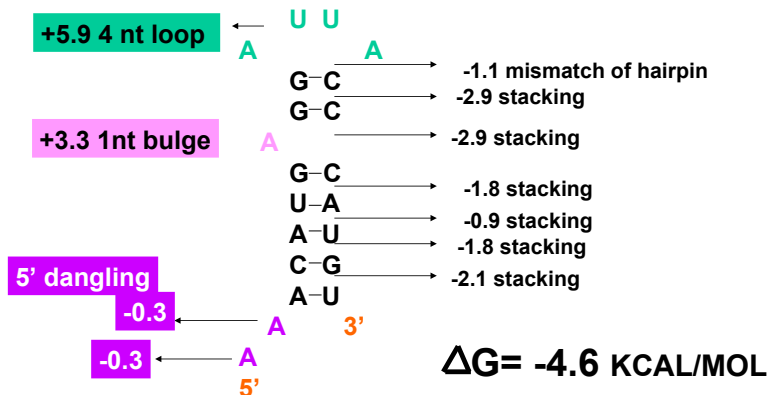Introduction
RNA folding by free energy minimization
develop a new structure abstraction

Free energy computation example
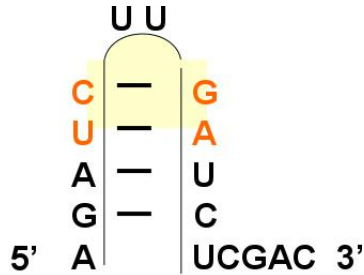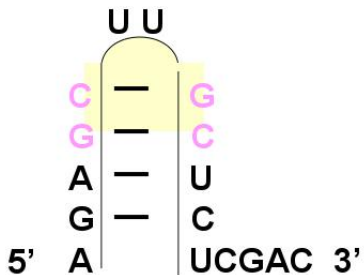
Example values:

GC	GC	GC	GC
AU	GC	CG	UA
-2.3	-2.9	-3.4	-2.1

Free energy computation example

+5.9 4 nt loop ←

U U

A

A

G–C → -1.1 mismatch of hairpin

G–C → -2.9 stacking

+3.3 1nt bulge

A → -2.9 stacking

G–C → -1.8 stacking

U–A → -0.9 stacking

A–U → -1.8 stacking

C–G → -2.1 stacking

A–U

5' dangling

-0.3 ← A   3'

**-0.3** ← A

5'

ΔG= -4.6 KCAL/MOL

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.

energy landscape
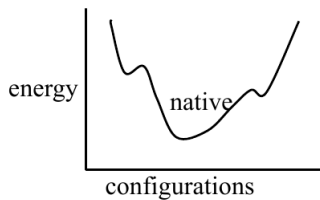
Introducing abstract shapes

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

- developed by Giegerich and Voss
- 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

CCCGUAGCUAGCGGUACGA

squiggle plot   *tree*   *abstract shape*
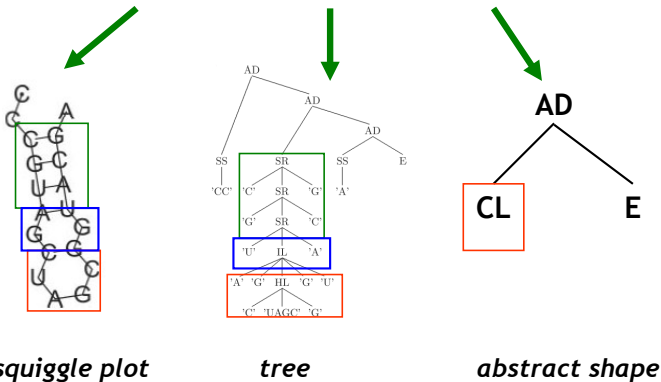
Abstract shape, Energy range: 5 kcal/mol

```
        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

Drawback of abstract shape

```
        AACUAAAACAAUUUUUGAAGAACAGUUUCUGUACUUCAUUGGUAUGUAGAGACUUC
 -9.00  ........................(((((((((((((.....)))))..))))))))..  []
-10.70  ..((...(((((((..(((((.(((((...)))).)))))..))).)))..)).....  []
```

Drawback of abstract shape

```
        AACUAAAACAAUUUUUGAAGAACAGUUUCUGUACUUCAUUGGUAUGUAGAGACUUC
-10.70  ..((...(((((((..(((((.(((((...)))).)))))..))).)))..)).....  []
 -7.70  ...(((((....)))))......(((((((((((((.....)))))..))))))))..  [][]
```

develop a new structure abstraction

The straightforward idea to overcome the position independence of the current available shape abstractions is
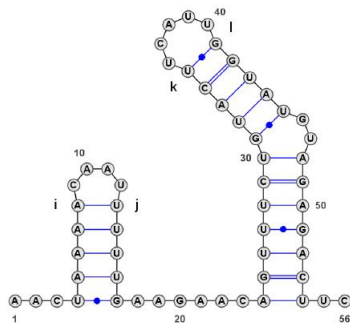
what keeps track of positions

- helical regions of hairpin loop
- helical regions of multiloops
- helical regions of stacked pairs, bulge and internal loops

which positions of the base pair will be tracked
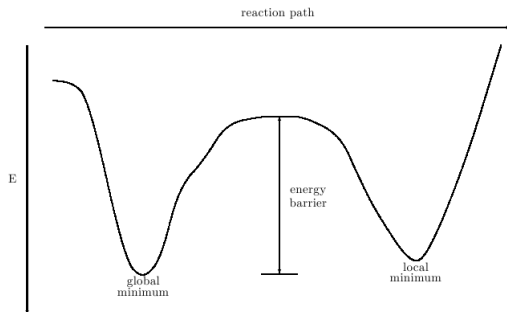
- i
- j
- (i,j)
- i+j/2

energy landscape

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]
```
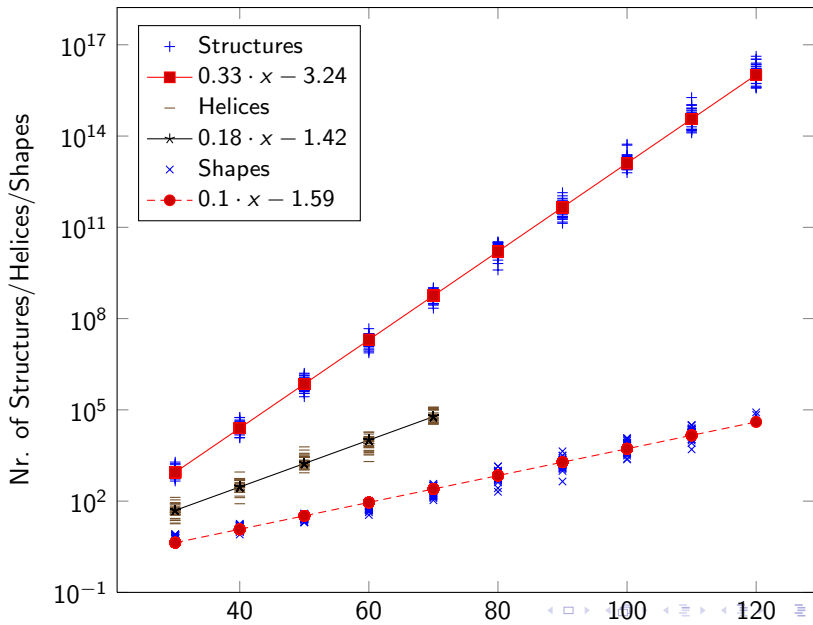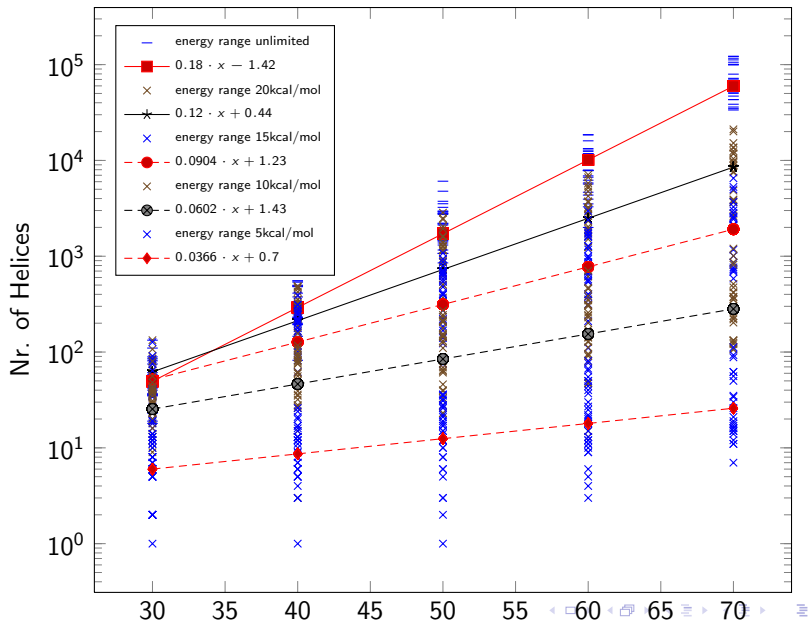
energy landscape

Possible problems

- do not preserve nesting of elements and meight lead to abstract shapes where a helix index occurs more than once $\rightarrow$ solved with different representation form

- abstracting from bulge and internal loops might be to rigorous $\rightarrow$ 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 $\rightarrow$ runtime problem

Growth of structure, helix and shape space

Growth of helix 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 ?