Helices indices of RNAs

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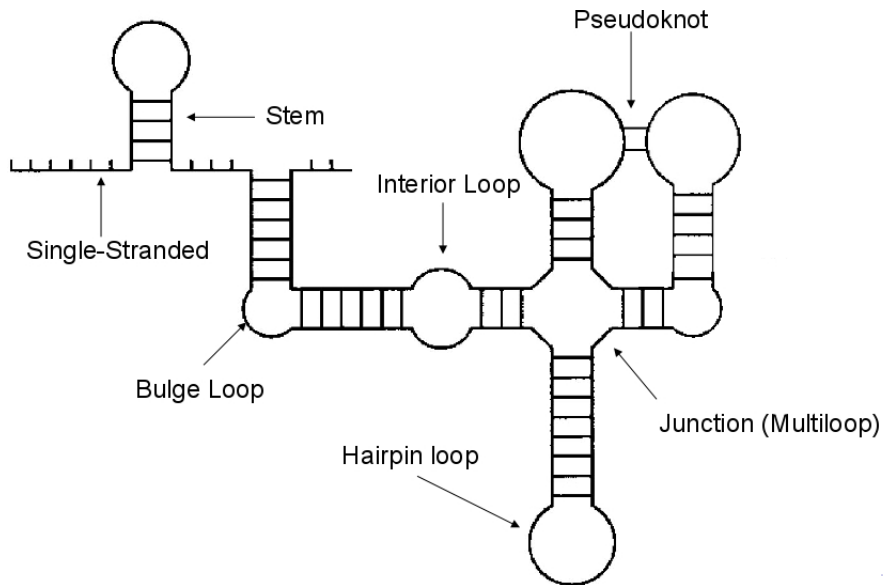
AG Experimentelle Bioinformatik (Cyanolab)
Institut für Biologie III
Universität Freiburg

March 10, 2011

Outline

- an overview of RNA secondary structure domains and a classical RNA secondary structure prediction algorithm
- introducing concept of abstract shapes
- basic idea about the helices indices
- outlook

Secondary structure elements of RNA

Pseudoknot

Stem

Interior Loop

Single-Stranded

Bulge Loop

Hairpin loop

Junction (Multiloop)

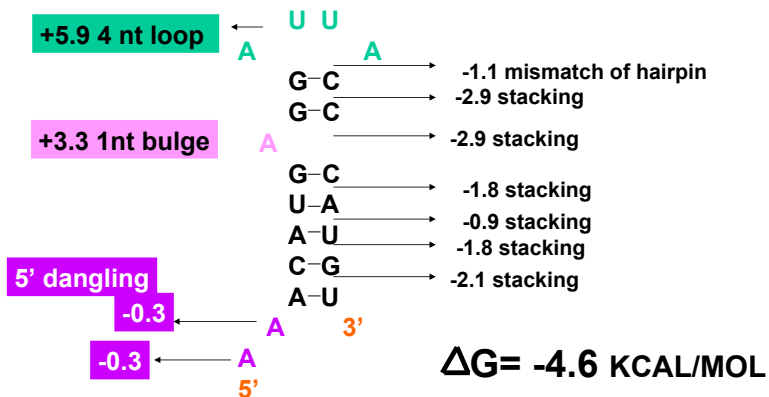# Classical secondary structure algorithms (Zuker 1981)

Simplifying Assumption for Structure Prediction

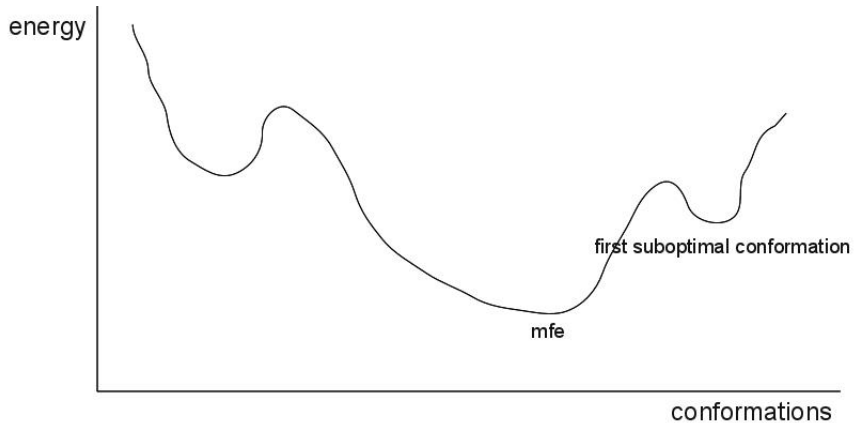- There are no knots (base pairs never cross).

facts about Zuker algorithm

- first described by Zuker and Stiegler in 1981
- idea: every secondary structure can be attached with an energy value according to some rules, the algorithm computes with a given RNA sequence the structure with the minimum free energy.
- the runtime is $O(n^3)$
- can get only one structure

Free energy computation example

U U
**+5.9 4 nt loop**   ←
A   A   → -1.1 mismatch of hairpin
G–C   → -2.9 stacking
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   **ΔG= -4.6** KCAL/MOL
                     5'

energy landscape

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
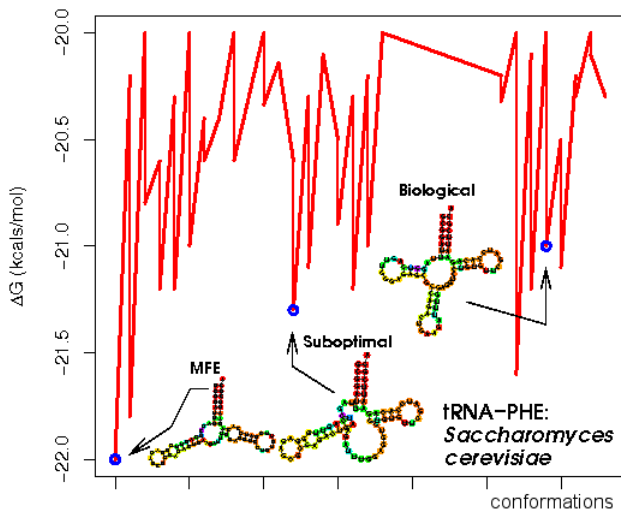
Introducing abstract shapes

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).

Abstract shape, energy range: 5 kcal/mol

(a) (b) (c)

(d)

Shape		
GGGCCCAUAGCUCAGUGGUAGAGUGCCUCCUUUGCAAGGAGGAUGCCCUGGGUUCGAAUCCCAGUGGGUCCA		
[]	(((((((((((((((((.((((.....(((((((...))))))).)))))))))))........))))))))).	-35.9 kcal/mol
[[][]]	(((((((.....((.((((.....(((((((...))))))).)))))((((.......))).))))))))).	-32.2 kcal/mol
[[][][]]	(((((...(((((.......)))).(((((((...))))))).....(((((.......))))).))))))).	-31.7 kcal/mol

Drawback of abstract shape

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

Drawback example 1 of abstract shape

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

Drawback example 2 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

the position of which secondary structure domains will be tracked

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

which positions of the vordefined structure domain will be tracked
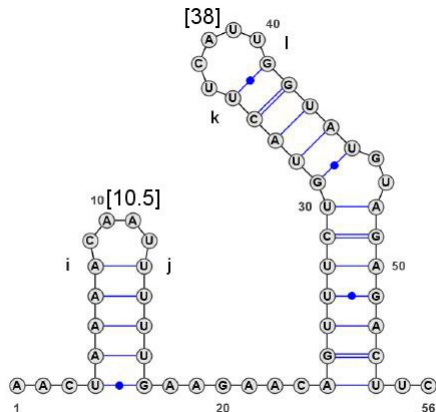
- i
- j
- (i,j)
- i+j/2

helices positions example

Figure: The structure is composed of two helices which are closed by hairpin loops (i,j) and (k,l), respectively. The positions are: i=8, j=13, k=35 and l=41. Thus, this structure would be abstracted to [10.5,38]
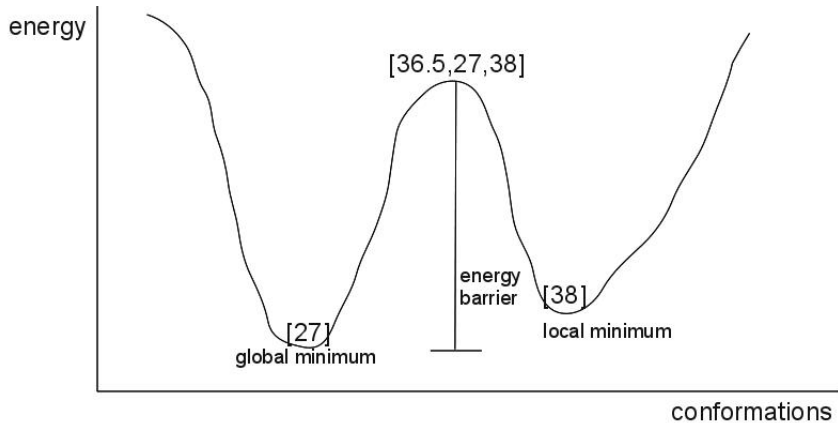
Output from the first trial version of helices indices

```
-10.7 ..((..(((((..(((((.((((...))))..)))))..)))..)))..))..... [27] *
 -9.0 ......................(((((((((((((.....)))))..)))))).. [38] *
 -6.7 ..((..((......))..))...(((((((((((((.....)))))..))))))).. [11.5,38]
 -6.6 ..((.(((....)))..))...(((((((((((((.....)))))..))))))).. [11,38]
 -6.4 .(((((........((((.((((...))))..)))))))))))..((....))).. [27,49.5]
 -6.2 ..((((((....))))..))...(((((((((((((.....)))))..))))))).. [10.5,38]
 -5.5 ..((....((((...)))..))...(((((((((((((.....)))))..))))))).. [13,38]
 -5.0 ..(((((((...))))...))...(((((((((((((.....)))))..))))))).. [10,38]
 -4.8 .(((((........((((.((((...)))).)))))))))))).....((....)) [27,52.5]
 -4.6 ..(((((....)))....))...(((((((((((((.....)))))..))))))).. [9.5,38]
 -4.2 .(((((........((((.((((...)))).)))))))))))).....((....)).. [27,51]
       ...
 -3.3 .(((((((....))).(((((.((((...)))).)))))...))).............. [9.5,27,22.5]
 -3.3 .(((((........((((.((((...)))).)))))))))))).....((....)). [27,51.5]
 -3.3 ..(((((.....))....))...(((((((((((((.....)))))..))))))).. [9,38]
 -3.3 ..(((.((....))).(((((.((((...)))).)))))........))....... [10.5,27,26]
 -3.3 ..(((.((.....))..))...(((((((((..((...)))..)))))))))).... [11,39]
 -3.2 ..((.((......))..))...((((((((......(((....))))))))))))).. [11.5,41.5]
 -3.1 ..((.((((.....))..))...(((((((((......(((....))))))))))))).. [11,41.5]
 -3.0 .(((((........((((.((((...)))).)))))))))))).....((....)). [27,52]
 -2.9 ..(((((((....))))..))...((((((((((..((...))..)))))))))))).. [10.5,39]
 -2.9 ..(((..........((((.....)))).((((((.....))))).)))...... [23,38,26]
 -2.9 ..(((.............((((....)))).((((((.....))))).)))...... [24.5,38,26]
 -2.8 .((.((((((...)))))((((.((((...)))).))))........))...... [10.5,27,24.5]
 -2.8 .((.((((...)))((((.((((...)))).)))))............))..... [10.5,27,27]
 -2.8 ..((...(((((((..((((......((((......)))))))..))))..)))).).... [29.5]
 -2.7 .((.((((...)))((((.((((...)))).)))))............)........ [9.5,27,24.5]
 -2.7 ..(((((((....))))..))...(((((((((......(((....))))))))))))).. [10.5,41.5]
 -2.7 ..(((...........((((((((.(((((...)))).))))).((......))..)).... [27,44,27]
 -2.4 ..(((((....)))((((.((((...)))).)))))............))..... [9.5,27,27]
 -2.3 .................((((..(((...)))(((((.....)))))........)))) [36.5,27,38] *
```
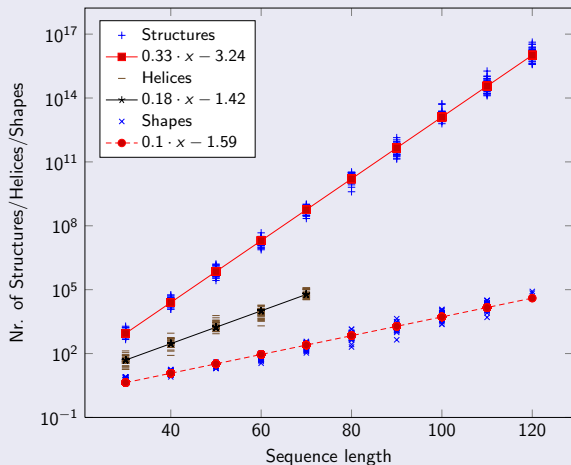
energy barrier

Possible problem

Although the helices space grows considerably slower than structure space, but it is still exponential
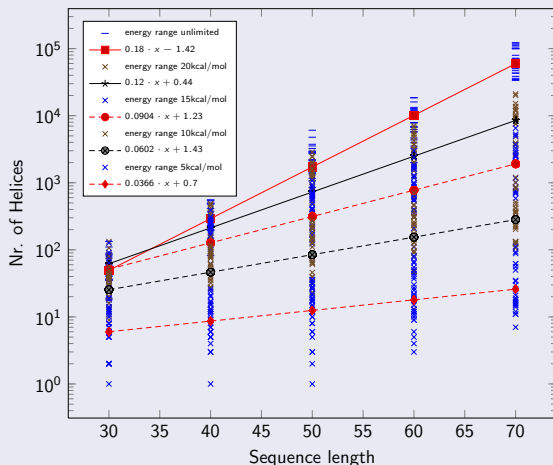

Growth of structure, helix and shape space

Solution of possible problem

One of the solution: we limit the helices space by setting an energy range on it

Growth of helix space

Outlook: Designing RNA class predictors

- Development of a new structure abstraction (helices index)
- Implementation of an algorithm based on the structure abstraction
- Evaluating the algorithm
- Designing RNA class predictors.

- Thanks a lot for your attention !
- Questions ?