Helices indices of RNAs

Jiabin Huang

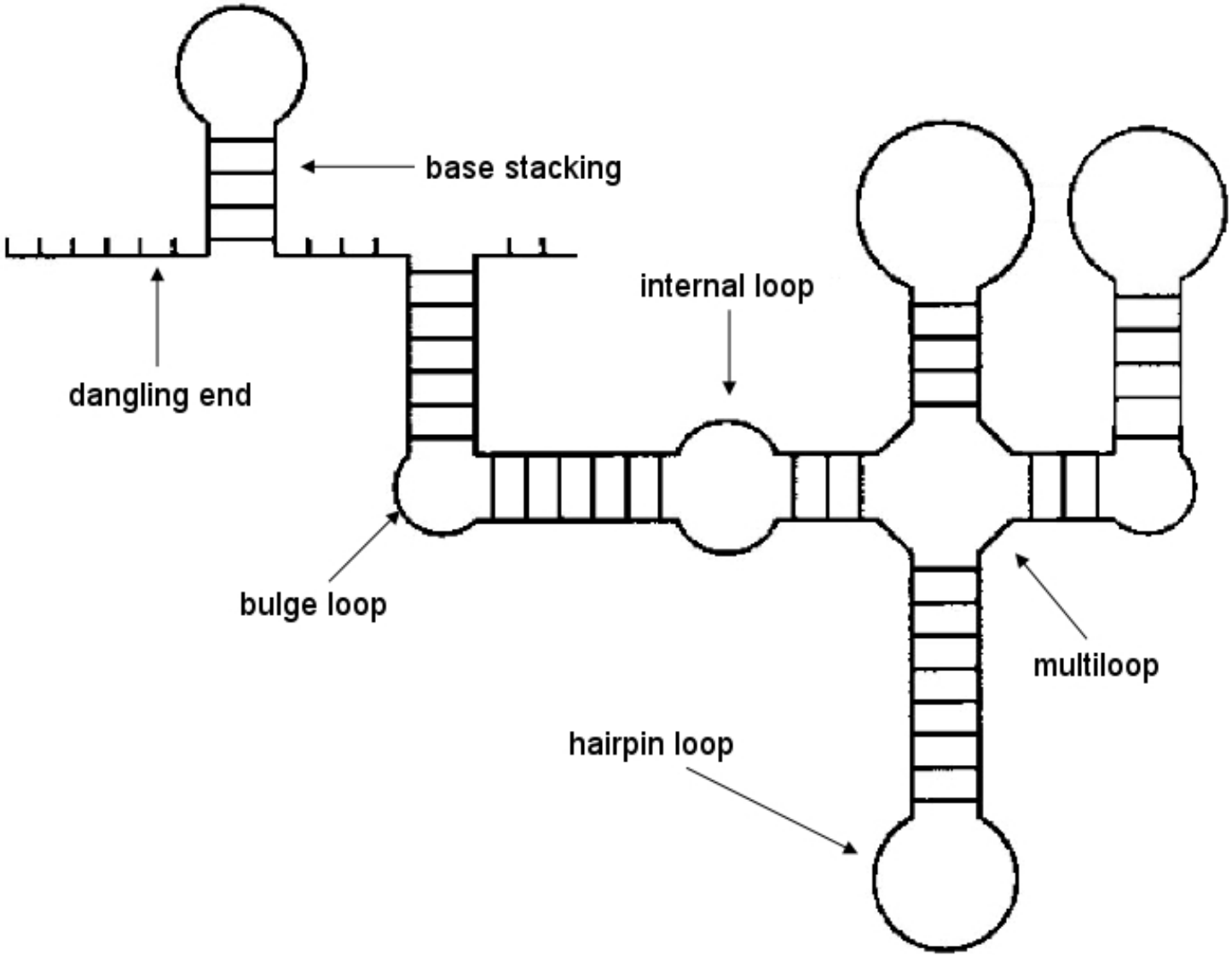
AG Experimentelle Bioinformatik (Cyanolab)
Institut für Biologie III
Universität Freiburg

March 11, 2011

Outline

- an overview of RNA secondary structure elements 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

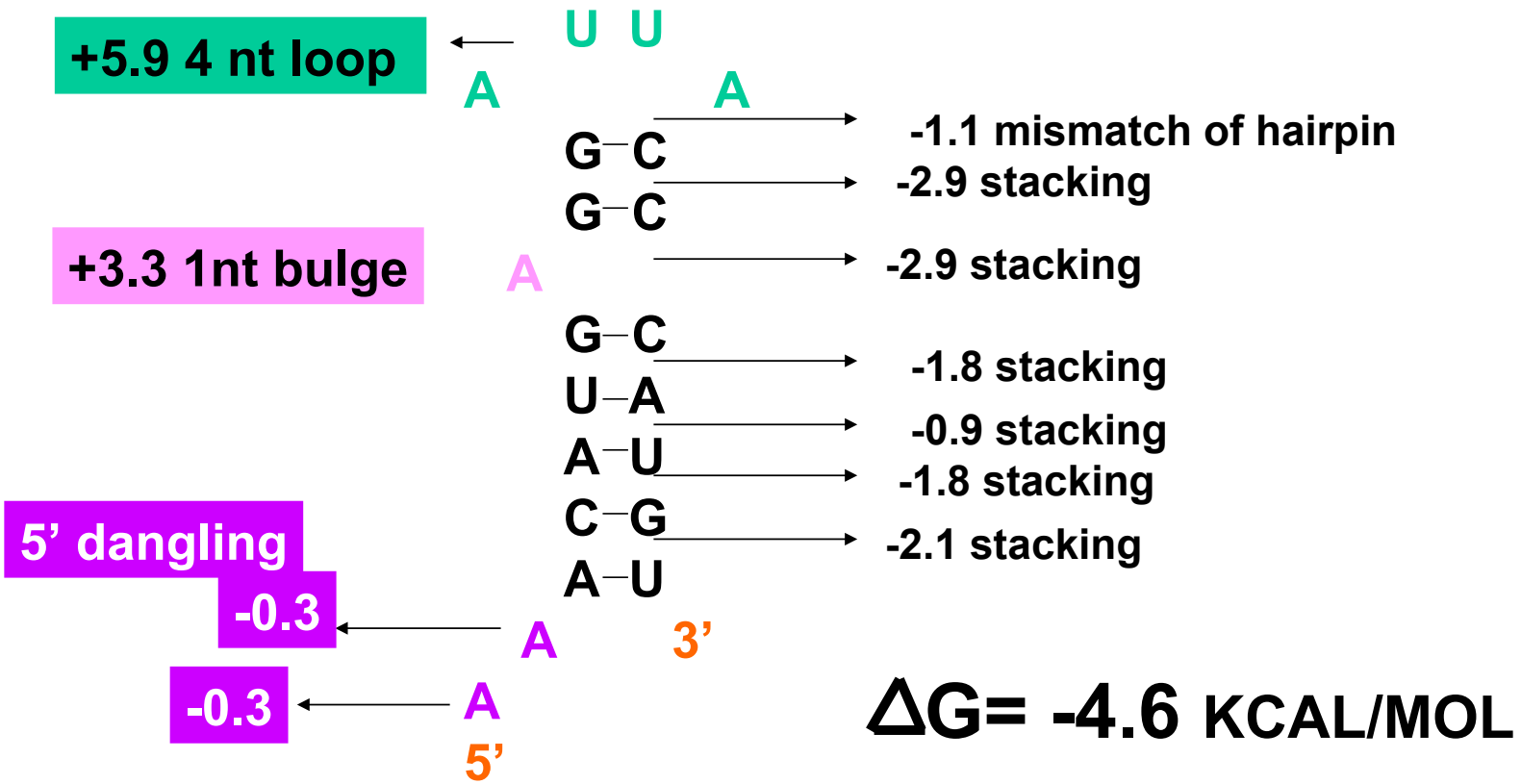


Classical secondary structure algorithms (Zuker 1981)

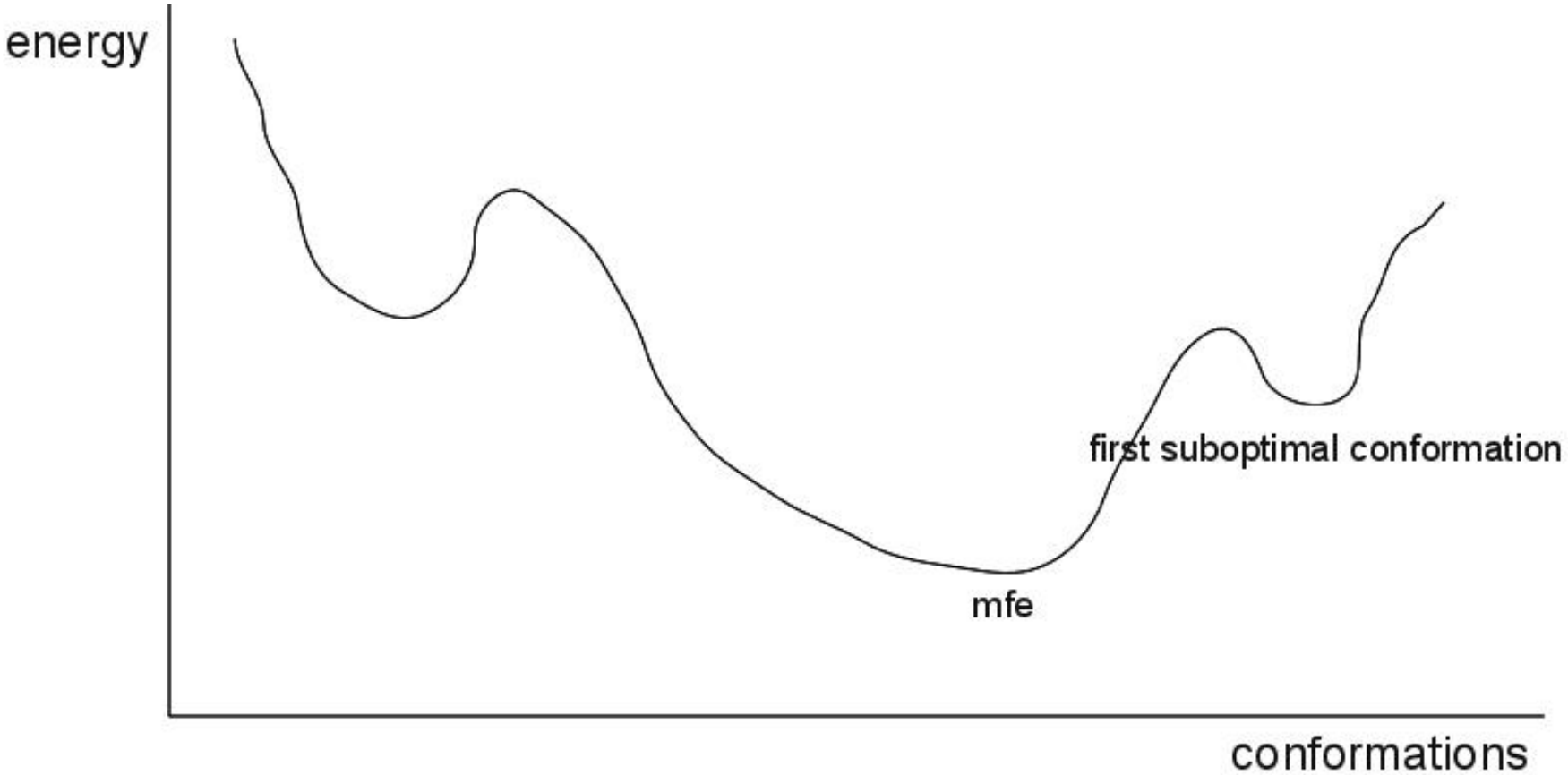
facts about Zuker algorithm

- first described by Zuker and Stiegler in 1981
- assumption: There are no knots (base pairs never cross).
- basic idea:
  1. a RNA sequence can be folded into many different secondary structure
  2. for every secondary structure, we can calculate a free energy value
  3. after that, the algorithm choose the structure with the minimum free energy
- the runtime is $O(n^3)$
- can get only one solution

Free energy computation example



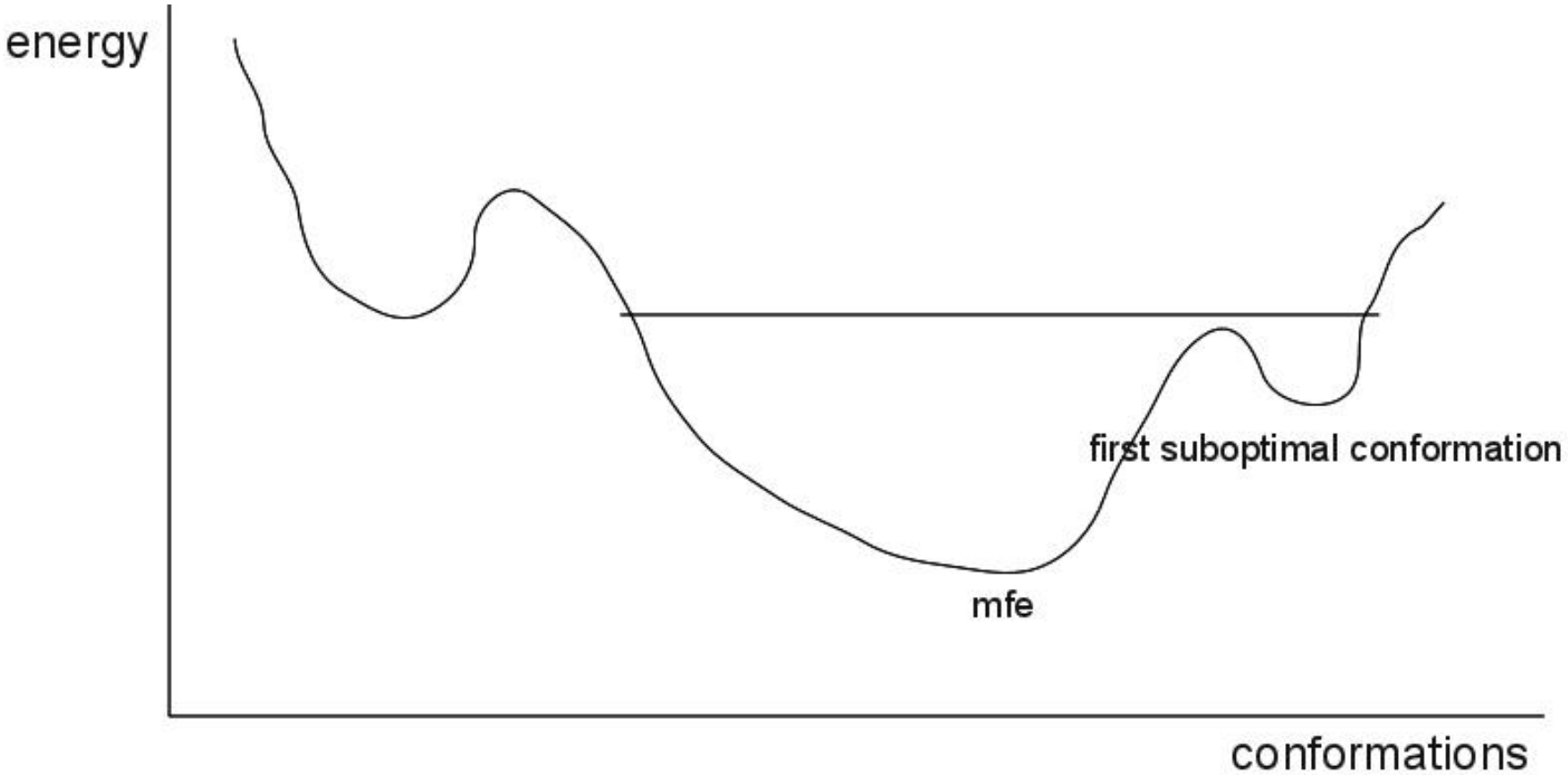
energy landscape



Suboptimal structures

- but the "true" structure is not always the one with the lowest predicted free energy.
- but it is not far away from the mfe and it is normally a local minimum
- one of solution: enumerate all suboptimal structures within a given energy range
- but the number of suboptimal structures grows exponentially with the energy range considered.

energy landscape setting a range

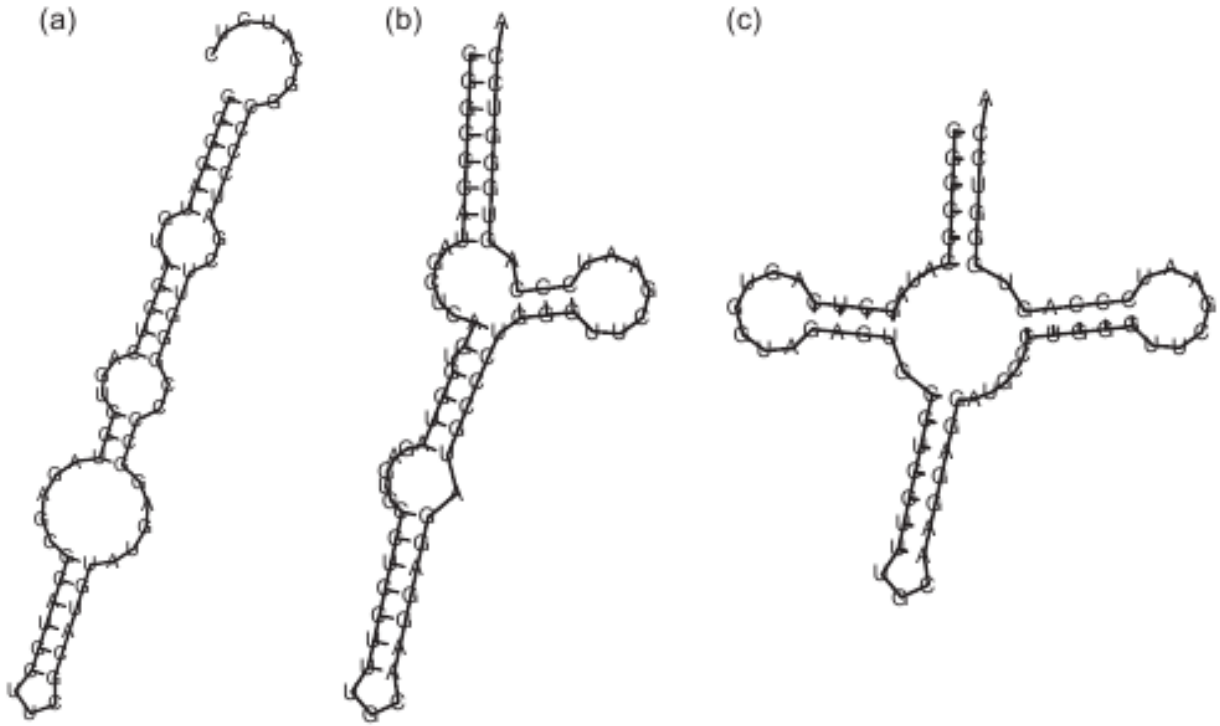


Introducing abstract shapes

Solution: further classify secondary structures within the lower energy range with different definition.

- abstract shapes is one approach in this direction
- developed by Giegerich and Voss
- intial idea:
  1. the user is usually only interested in structures that show fundamental differences
  2. small changes, such as additional base pairs or changing bulge loops are of minor significance
- central to this approach: do not care about all details of the structures and abstract from lengths of structural elements or from some elements in total
- each shape has a representative structure called shrep (with minimum free energy within the shape class)

Abstract shape, energy range: 5 kcal/mol



(d)

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

Drawback of abstract shape

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

Drawback example of abstract shapes

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

develop a new structure abstraction: helices indices

The straightforward idea to overcome the position independence of the current available shape abstractions is to develop a new structure abstraction that includes the information of positions of helices

Which element we should take to record the positions?

- hairpin loop
- multiloop
- bulge or internal loops
- any combinations of them

Which position of elements we should take?

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

helices positions example

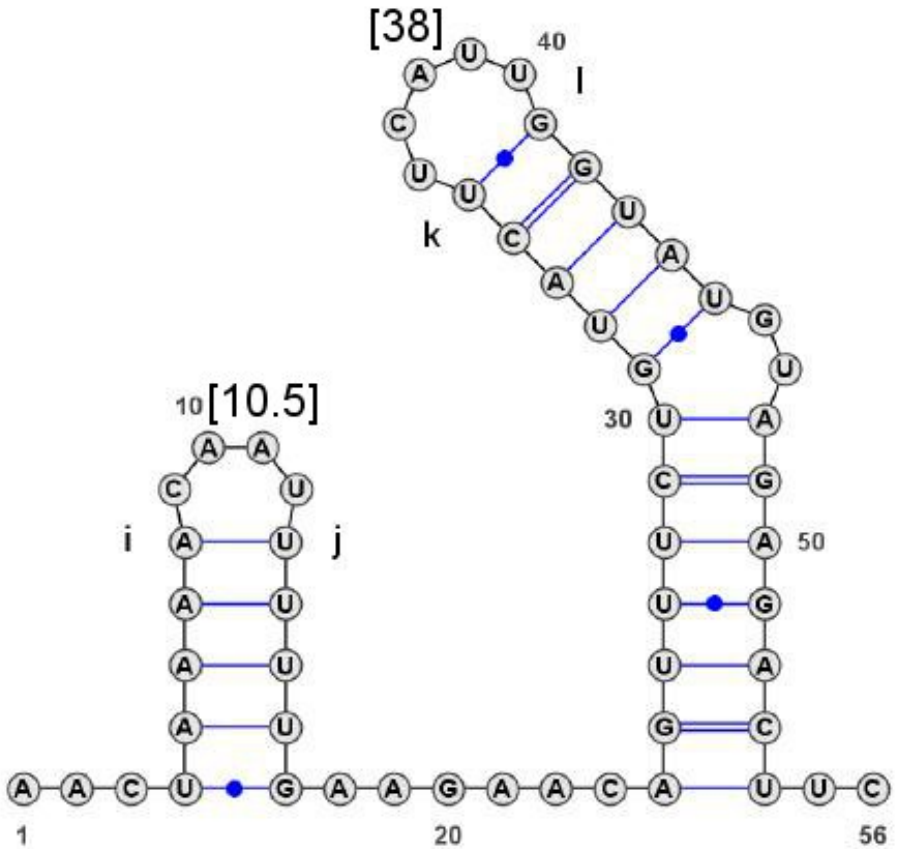


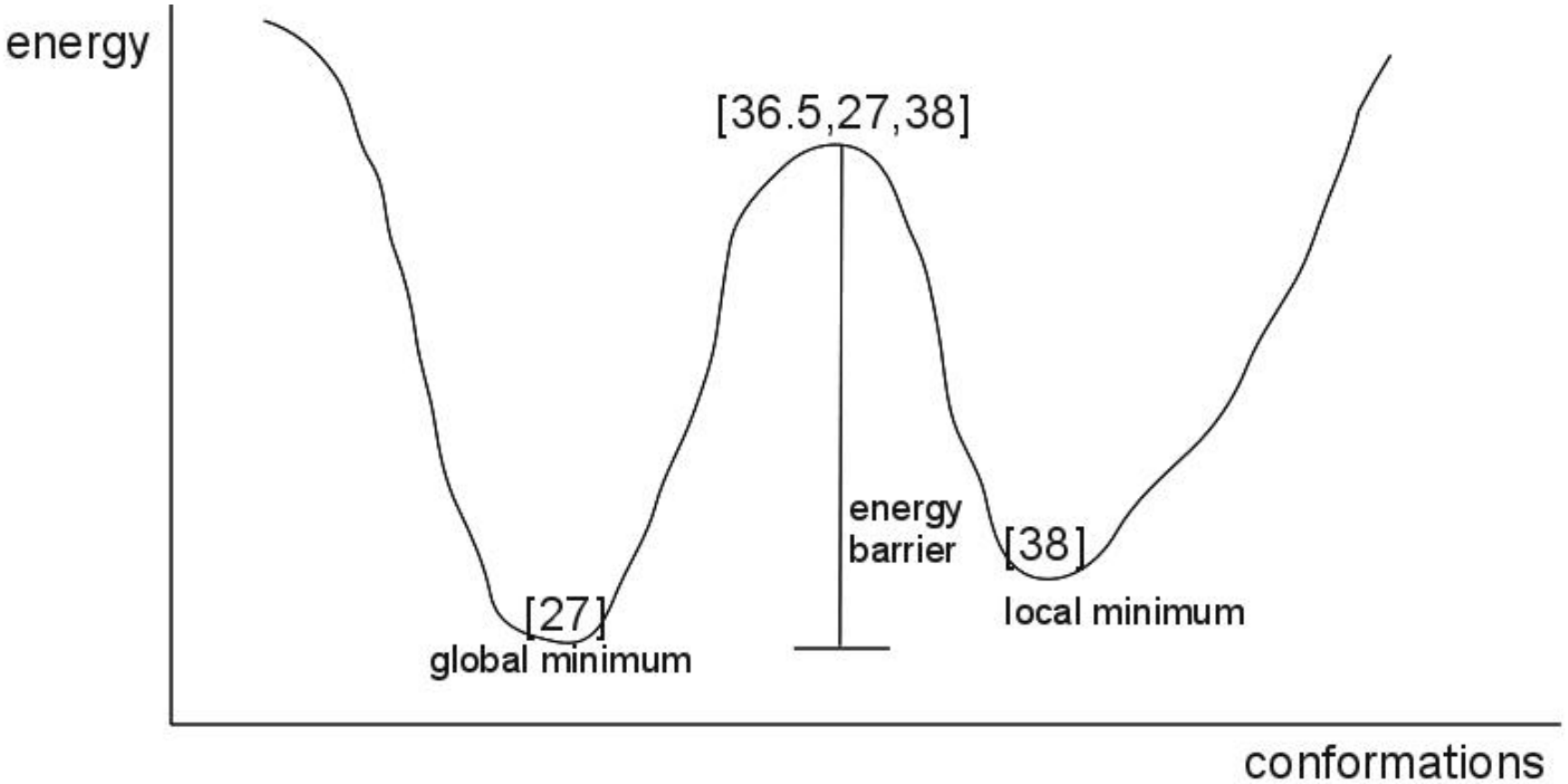
Figure: The structure is composed of two helices which are closed by hairpin loops (i,j) and (k,l), respectively. The positions are: i=8, j=13, k=35 and l=41. Thus, this structure would be abstracted to [10.5,38]

Output from the first trial version of helices indices

Helices shape, Energy range: 5 kcal/mol

```
-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] *
```

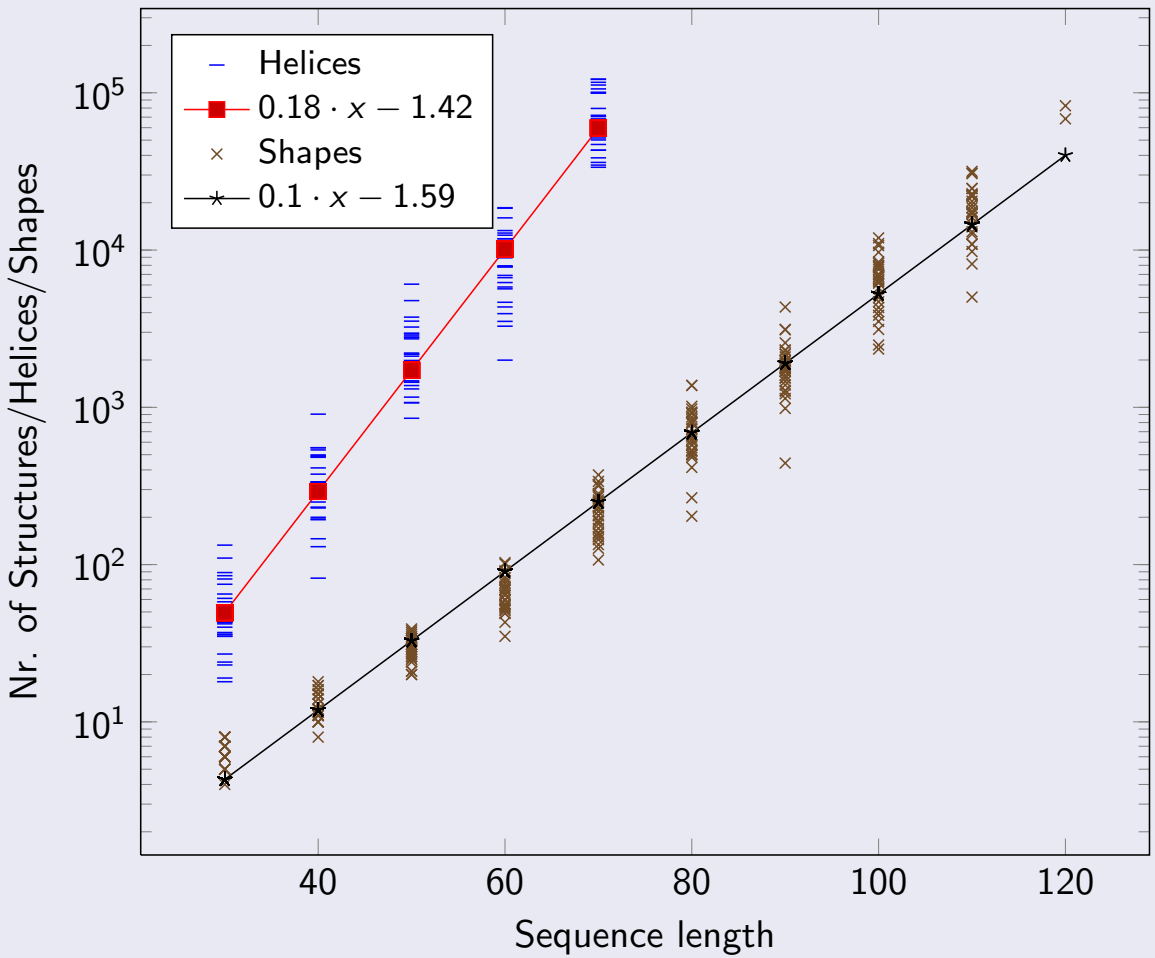
energy barrier



Possible problem

Although both the abstract shapes and helices indices space grow exponentially with sequence length and suboptimal energy range, the helices indices space grows considerably faster.

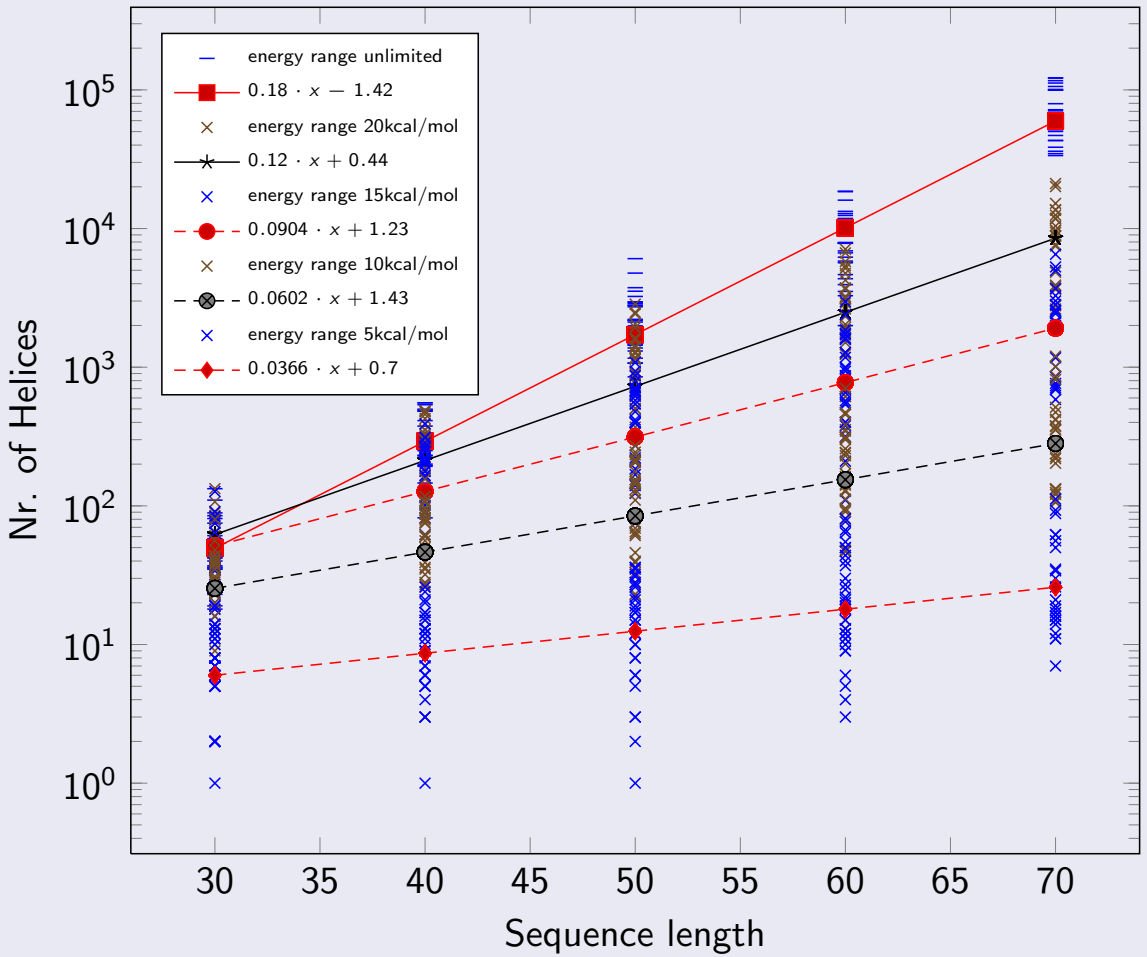
comparison of helices indices space with abstract shapes space



Solution of possible problem

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

comparison of helices indices space setting different energy ranges



Outlook

- develop a new structure abstraction (helices indices)
- implement a software with the ADP framework in C++ programming language based on the idea
- evaluate the algorithm comparing with other programs
- design a RNA class predictors

End

- Thanks a lot for your attention !
- Questions ?