Kinwalker - Kinetic Backtracking of RNA folding

Michael Geis

Bioinformatics Group, Dept. of Computer Science
University of Leipzig

TBI Winterseminar Bled 2007

Essential components in simulating RNA folding:

Select a set of states **Model state transitions**

Main issue: How to calculate energy barriers between states.

Calculate the suboptimal structures

For 1 <= i <= j <= n consider as states the lowest energy structures on a subsequence (i,j) given that bases i and j pair (extremes).

-	-	-	-	620	450	INF	310	INF	520	INF	INF	INF	INF	INF	-640	INF	INF	-60	-10	20	120	139	INF	160	INF	-150	INF	INF	INF	INF	INF	-1970
-	-	-	-	-	INF	560	INF	530	INF	530	230	450	240	-420	INF	150	190	INF	INF	INF	INF	INF	310	INF	40	INF	-430	-760	-650	-1010	-1720	INF
-	-	-	-	-	-	570	INF	470	INF	470	INF	390	-180	INF	INF	INF	100	INF	INF	INF	INF	INF	INF	INF	-50	INF	-520	INF	-860	-1570	-1210	INF
-	-	-	-	-	-	-	INF	420	INF	340	INF	150	220	INF	INF	INF	20	INF	INF	INF	INF	INF	INF	INF	-90	INF	-560	INF	-1240	-880	-770	INF
-	-	-	-	-	-	-	-	620	INF	480	360	430	370	140	INF	230	50	INF	INF	INF	INF	INF	240	INF	-80	INF	-480	-1030	-670	-560	-520	INF
-	-	-	-	-	-	-	-	-	INF	490	490	460	270	-60	INF	110	330	INF	INF	INF	INF	INF	200	INF	-120	INF	-860	-610	-510	-470	-430	INF
-	-	-	-	-	-	-	-	-	-	INF	INF	INF	INF	INF	320	INF	INF	70	310	310	520	330	INF	20	INF	-890	INF	INF	INF	INF	INF	-350
-	-	-	-	-	-	-	-	-	-	-	INF	40	400	INF	INF	INF	410	INF	INF	INF	INF	INF	INF	INF	-640	INF	-170	INF	-90	-50	-10	INF
-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	270	INF	INF	260	510	420	320	250	INF	-500	INF	20	INF	INF	INF	INF	INF	-10
-	-	-	-	-	-	-	-	-	-	-	-	-	620	490	INF	240	510	INF	INF	INF	INF	INF	-360	INF	120	INF	200	240	280	299	310	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	450	INF	INF	460	410	150	210	-260	INF	250	INF	330	INF	INF	INF	INF	INF	-31
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	INF	INF	INF	-130	120	INF	450	INF	530	INF	INF	INF	INF	INF	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	420	100	80	180	490	INF	570	INF	590	INF	INF	INF	INF	INF	-90
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	430	410	390	540	390	INF	670	INF	400	INF	INF	INF	INF	INF	-130
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	520	570	INF	630	INF	50	INF	INF	INF	INF	INF	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	INF	INF	INF	270	INF	410	INF	390	330	160	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	440	560	INF	480	INF	620	INF	INF	INF	INF	INF	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	620	540	INF	530	INF	210	INF	INF	INF	INF	INF	-830
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	460	INF	410	INF	370	60	-490	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	480	INF	410	INF	390	-160	-100	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	360	INF	340	INF	170	230	230	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	620	INF	480	380	440	440	480	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	510	500	490	530	500	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	INF	INF	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	620	480	530	510	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	INF	INF	280
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	620	530	INF
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	INF	430
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

–	–	–	–	620	450	–	310	–	520	–	–	–	–	–	−640	–	–	−60	−10	20	120	139	–	160	–	−150	–	–	–	–	–	−1970
–	–	–	–	–	–	560	–	530	–	530	230	450	240	−420	–	150	190	–	–	–	–	–	310	–	40	–	−430	−760	−650	−1010	−1720	–
–	–	–	–	–	–	570	–	470	–	470	–	390	−180	–	–	–	100	–	–	–	–	–	–	–	−50	–	−520	–	−860	−1570	−1210	–
–	–	–	–	–	–	–	–	420	–	340	–	150	220	–	–	–	20	–	–	–	–	–	–	–	−90	–	−560	–	−1240	−880	−770	–
–	–	–	–	–	–	–	–	620	–	480	360	430	370	140	–	230	50	–	–	–	–	–	240	–	−80	–	−480	−1030	−670	−560	−520	–
–	–	–	–	–	–	–	–	–	–	490	490	460	270	−60	–	110	330	–	–	–	–	–	200	–	−120	–	−860	−610	−510	−470	−430	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	320	–	–	70	310	310	520	330	–	20	–	−890	–	–	–	–	–	−350
–	–	–	–	–	–	–	–	–	–	–	–	40	400	–	–	–	410	–	–	–	–	–	–	–	−640	–	−170	–	−90	−50	−10	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	270	–	–	260	510	420	320	250	–	−500	–	20	–	–	–	–	–	−10
–	–	–	–	–	–	–	–	–	–	–	–	–	620	490	–	240	510	–	–	–	–	–	−360	–	120	–	200	240	280	299	310	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	450	–	–	460	410	150	210	−260	–	250	–	330	–	–	–	–	–	−31
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−130	120	–	450	–	530	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	420	100	80	180	490	–	570	–	590	–	–	–	–	–	−90
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	430	410	390	540	390	–	670	–	400	–	–	–	–	–	−130
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	520	570	–	630	–	50	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	270	–	410	–	390	330	160	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	440	560	–	480	–	620	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	620	540	–	530	–	210	–	–	–	–	–	−830
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	460	–	410	–	370	60	−490	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	480	–	410	–	390	−160	−100	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	360	–	340	–	170	230	230	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	620	–	480	380	440	440	480	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	510	500	490	530	500	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	620	480	530	510	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	280
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	620	530	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	430
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Order the states

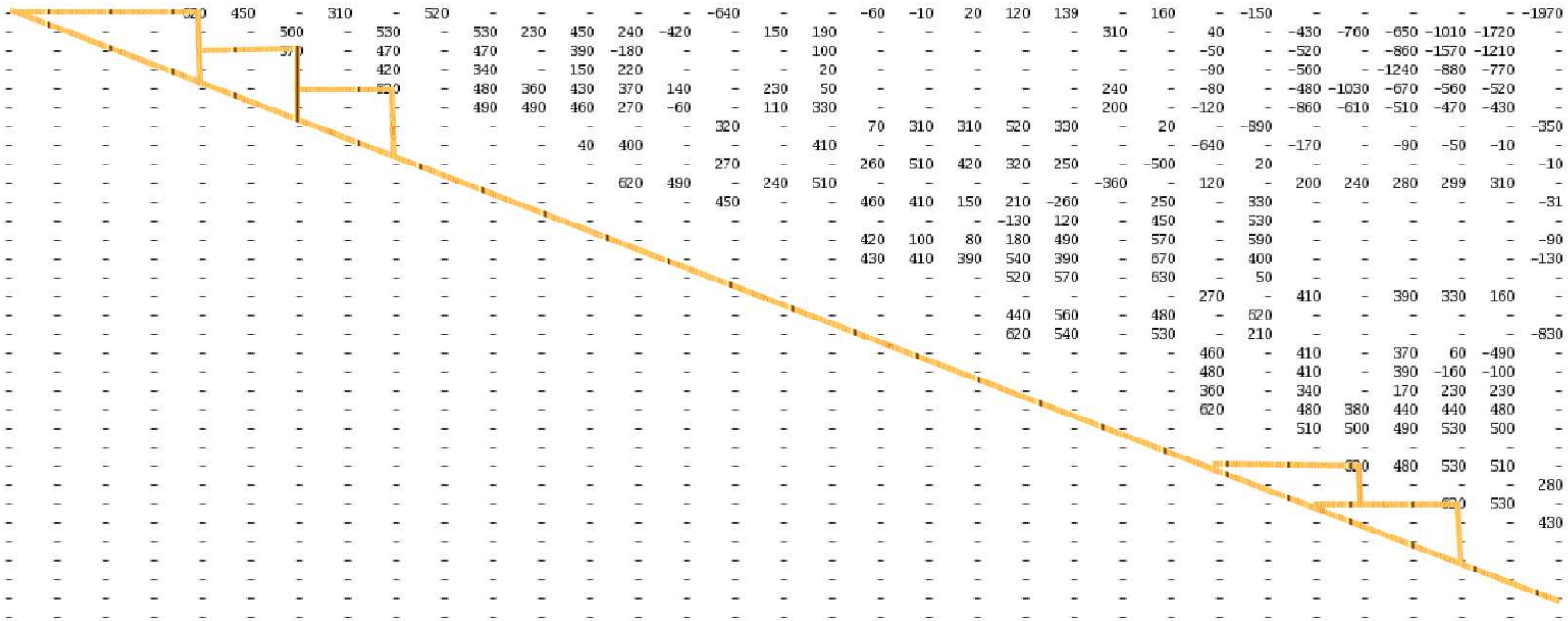
Rank extremes by

1. diagonal
2. distance from edge of matrix
3. 5’ before 3’

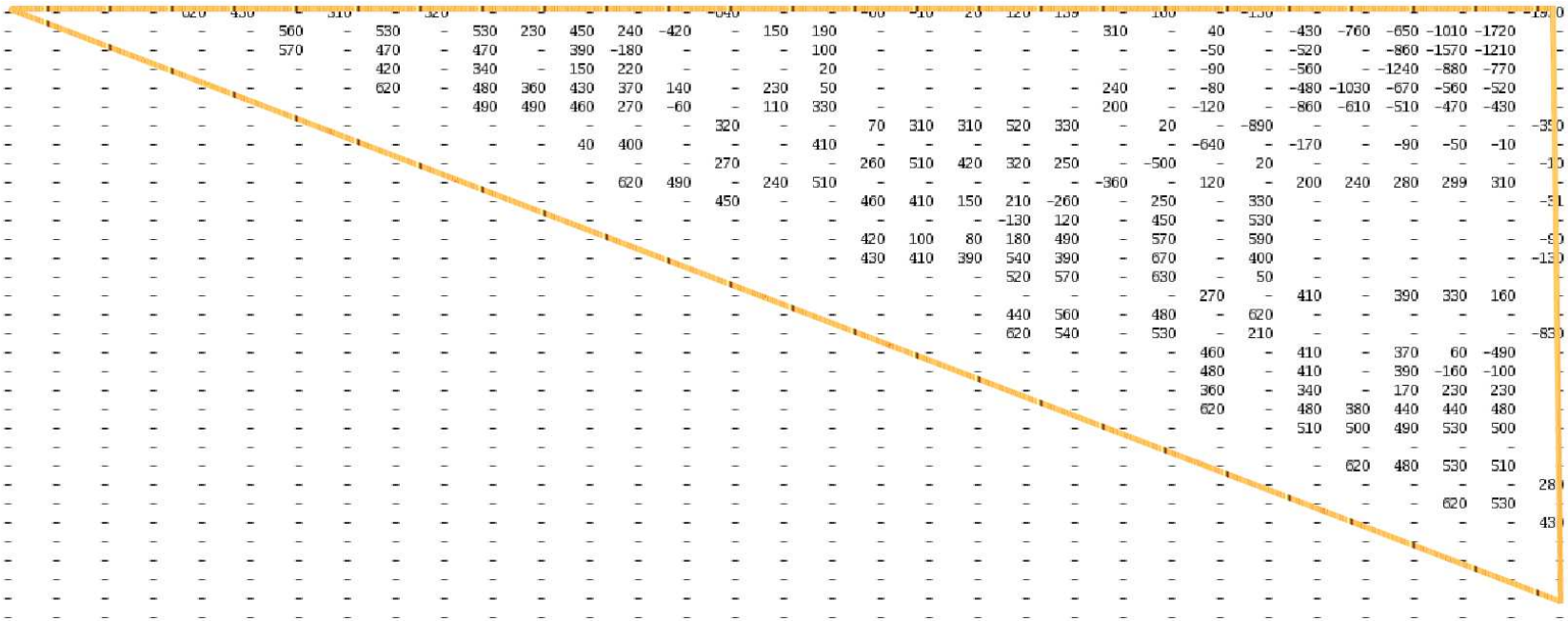
Algorithm

$saddleheight = 0$
$Front\ F = initial\ structure$
$while(F != mfE)$
    increment saddleheight
    for x in extremes
        calculate saddle point S between x and F
        $if(E(S) - E(F) < saddleheight)$
            extend F by x

Extend Front



Final Front



Conflict

A substructure (i,j) determines how the bases in that subsequence pair. Once (i,j) is part of the front, and a pair (k,l) overlapping with (i,j) is added, they may demand incompatible base pairs in the overlapping region.

Conflict Resolution

1. Add only those basepairs of a conflicting substructure that don't conflict with the current structure.

2. Replace further base pairs if it improves energy.

Saddle Height

To go from conformation A to B, there are $(A \Delta B)!$ direct paths.

The best path does not have to be direct.

The saddle height is the highest energy of the lowest path:

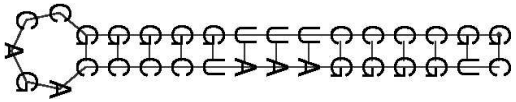
$$S = min_{P:\, A->B} max_{x \in P} E(x)$$

Heuristic for Saddle Points: Morgan-Higgs(1998)

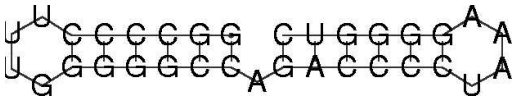
1. Rank the elements in $B \backslash A$ by conflict
2. For each x in $B \backslash A$
   i) Remove the base pairs in A that conflict with x
   ii) Add x
   iii) Add all other elements in $B \backslash A$ that can be added now
   iv) Record the energies of the traversed states
3. Take the highest recorded energy as saddle energy

Toy problem

MfE



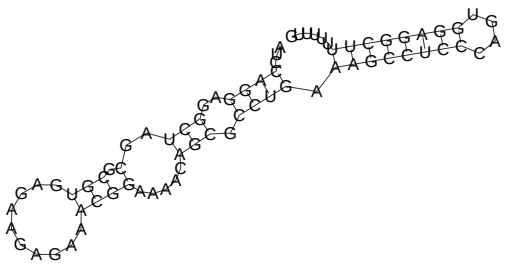
Metastable



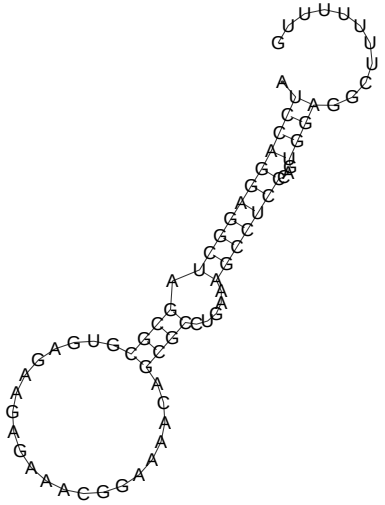
```
.............................
...((....))..................
...((....))..........((....)).....
..(((....)))........(((....)))...
(((((....)))))....(((....)))...
(((((....)))))...(((....))))..
(((((....)))))..((((....)))))..
(((((....)))))).(((((....)))))
```

E. coli Phenylalanin

MfE

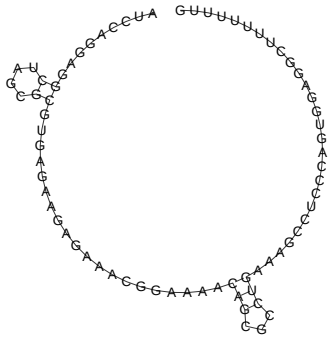


Metastable

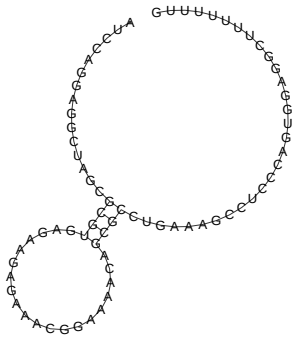


Folding Trajectory

State 1

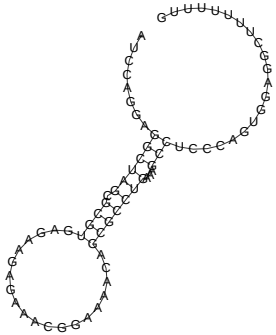


State 2

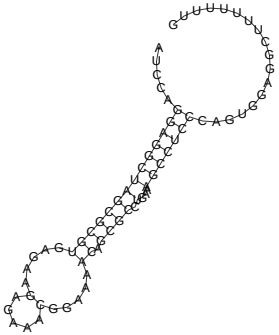


State 3

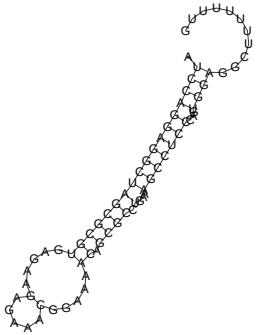
State 4



State 5

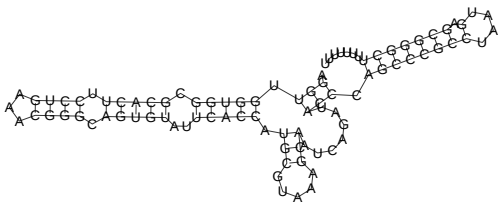


State 6

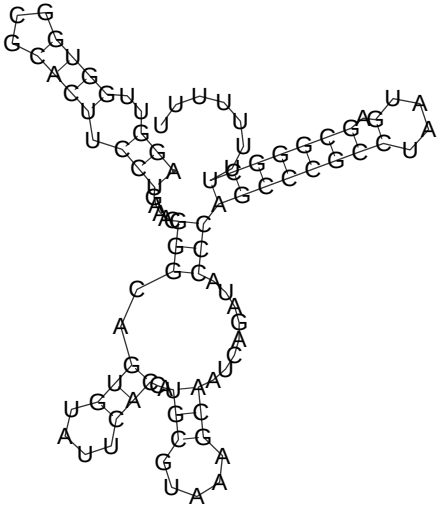


E. coli Tryptophane

MfE



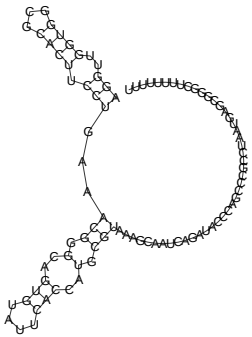
Metastable



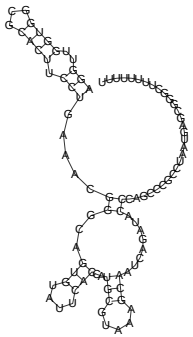
Folding Trajectory

State 1

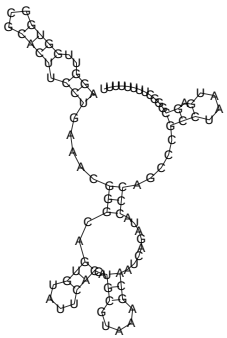
State 2



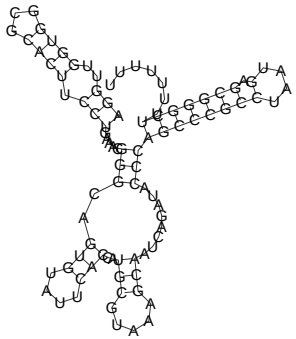
State 3



State 4



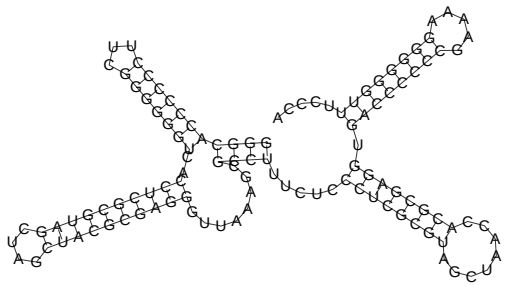
State 5



SV11

MfE

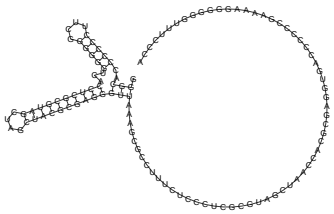
Metastable



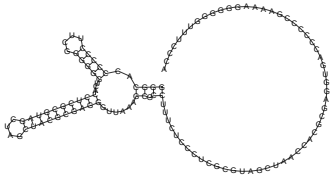
Folding Trajectory

State 1

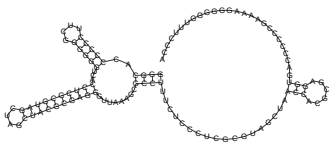
State 2



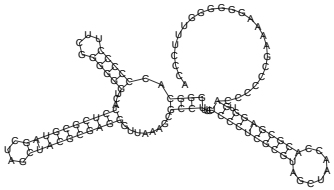
State 3



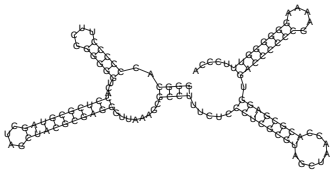
State 4



State 5

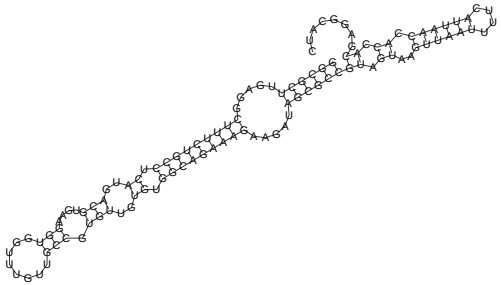


State 6

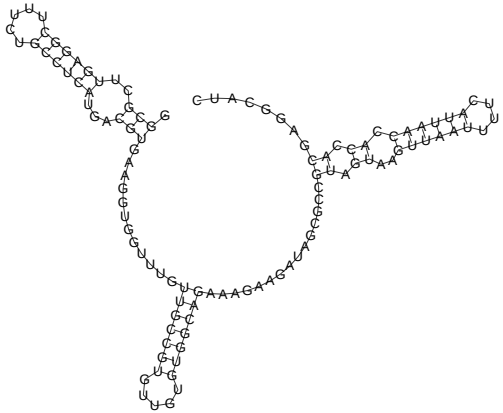


HOK

MfE



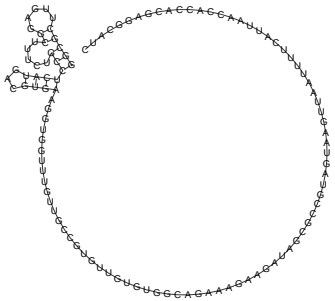
Metastable



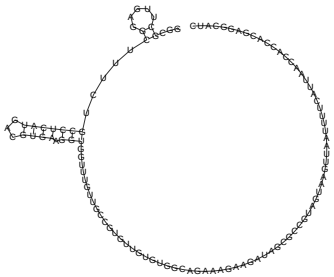
Folding Trajectory I

State 1

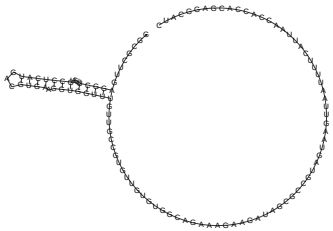
State 2



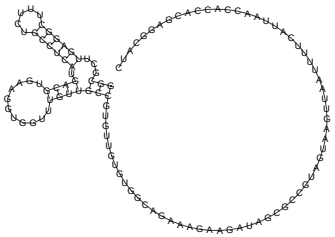
State 3



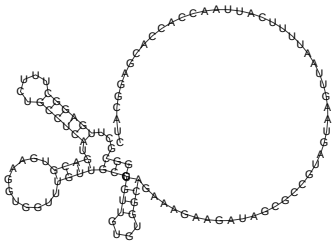
State 4



State 5



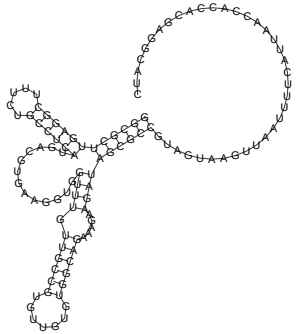
State 6



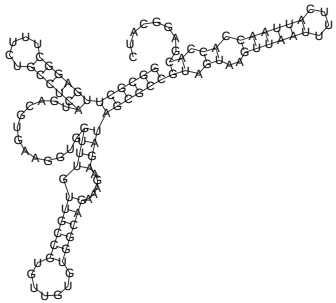
Folding Trajectory II

State 7

State 8



State 9



THANK YOU!

Caro Thurner, Christoph Flamm, Peter Stadler