k	1	2	3	$\sum_{k=1}^{3} P_k$
P_k				

1 Multiple optimal solution

Depending on how the traceback is implemented, only one of the origins with identical scores is investigated. However there can be multiple traces which result in the optimal solution. To obtain random optimal solutions, I would select a random origin if there are multiple origins with the same score. This can be easily implemented by first computing all origins and then selecting one at random. But this comes at a higher cost as we always have to compute the fourth case which increases the computations time.
To find all optimal solutions, in the worst we get a score of 0 in the upper right corner. The traceback can be saved in a traceback matrix. For every element in the matrix, the algorithm can origin in all four cases. There are $\frac{n}{2}$ elements, that can come from 3 origins(cases 1-3) plus n(case4). The runtime for all tracebacks is therefore $O(\frac{n}{2} * (3 + n))$

2 Analysis of RNA sequences

2.1

The scores of AU and GC in the range of [1,5,10] were tested. The score of gu stayed 0. The minimum loop length was also tested in the interval [1,5,10]. The results are shown in the figure1-5. Most obvious is the result when the minimum loop length is compared, like in 5 and 4. For the high minimum loop length in 4 the secondary structure only forms base pairs with bases far away, which leads to a worm like structure. A different picture is shown in 5, where the structure is very nested. In 1 the score for gc and in 3 the score for au is increased. In 1 an side arm is established at the beginning, whereas in 3 the structure has a side arm the is in the middle of the sequence. Most valid seems the structure shown in 2, where gc and au both got the value 5. Of the shown results this one seems to be the most valid, as there are many base pairs and the structure seems the most compact.
Out of interest a extensive search for the best structure was implemented, where the predicted one was compared to the original structure and the identity was used as score. The best result was achieved with au= 1 gc= 3 gu= 2 and a minimum loop length of 1. But this makes biologically no sense as au should have a higher score than gu. In my opinion the best result was achieved with au=5, gc=5, gu=0 and a minimum length of 5.

2.2

When we compare the best prediction by nussinov in 2 to the result in 8, the results look comparable, but the indices of the are shifted. The arms of the structure in are comparable in their length. Compared to the experimentally derived structure in 6, both structures fail in having a fourth side arm. The main multiloop at G10 in 8 fits better to the original structure than in 2. Also the hairpin loop at C60 in 8 fits slightly to the original structure.
The mFold performs better, as it is based on Zuker-Stiegler algorithm [1]. It is based on the observation that only k-loops that are not stacked pairs contribute to the free Enegry ΔG of a structure. For a given sequence the structure with all its k-loops is computed and the energy of the single k-loops are summed up. Instead of maximizing the the score of a structure by maximizing the most stacked pairs like in nussinov, the algorithm yields to find the composition of k-loops, that induce the least energy. This makes more sense in a biological way, therefore the algorithm performs better in predicting the correct structure.

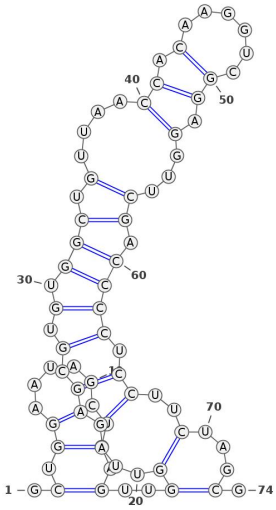

Abbildung 1: loop= 5, au= 1, gc=10

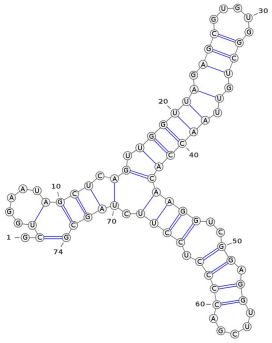

Abbildung 2: loop= 5, au= 5, gc=5

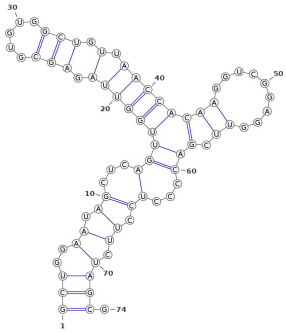

Abbildung 3: loop= 5, au= 10, gc=1

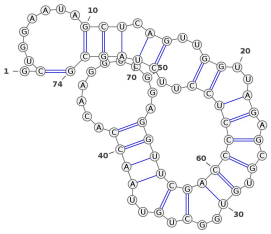

Abbildung 4: loop= 1, au= 5, gc=5

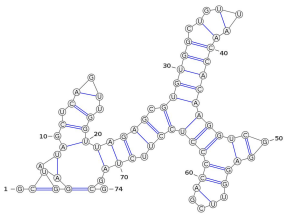

Abbildung 5: loop= 10, au= 5, gc=5

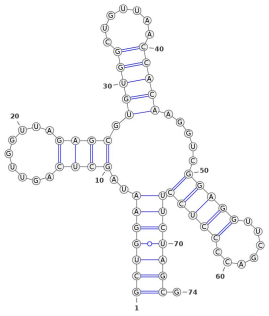

Abbildung 6: real structure

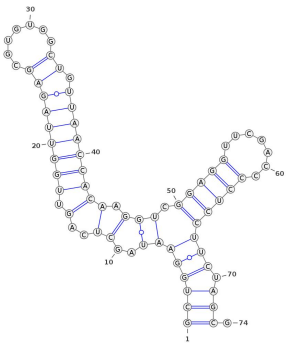

Abbildung 7: structure after extensive search with the identity to original structure as score.

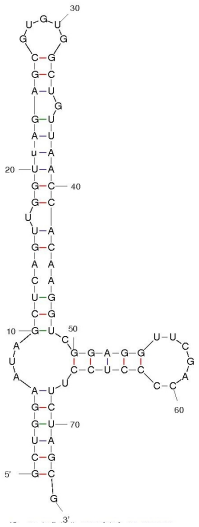

Abbildung 8: Prediction by mFold

3

Tabelle 1: k-loops of structure

Base pair	Value of k	Size	Bases being accesible	Name of secondary structure
3-51	2	0	4,50	stacked pair
4-50	2	0	5,49	stacked pair
5-49	4	3	6,7,21,22,23,46,47,48	multiloop
7-21	2	0	8,20	stacked pair
8-20	2	3	9,10,11,18,19,20	interior
11-18	2	3	12,13,14,16,17	interior
14-16	2	1	15	hairpin
23-45	2	1	24,25,44	bulge
25-44	2	3	26,27,28,42,43	interior
28-42	2	1	29,40,41	bulge
29-40	3	0	30,34,35,39	multiloop
30-34	1	3	31,32,33	hairpin
35-39	1	3	36,37,38	hairpin
46-47	0	0	0	null loop

Literatur

[1] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3406–3415, 2003. [Online]. Available: http://dx.doi.org/10.1093/nar/gkg595