

Applied stochastic RNA folding:

flexible move types, new parameters, and a modern framework



`fuzzyfold`: a high-performance framework for
stochastic RNA folding kinetics

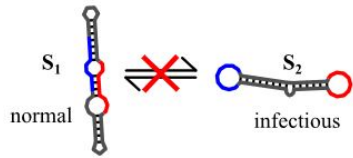
Stefan Badelt*¹

¹Department of Theoretical Chemistry, University of Vienna, Austria

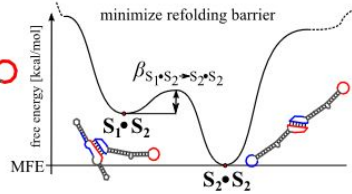
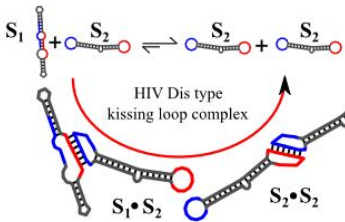
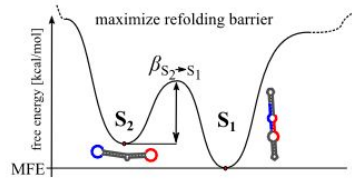
A personal history of secondary structure folding kinetics

Prion-like RNA design

Requirements for an RNA prion



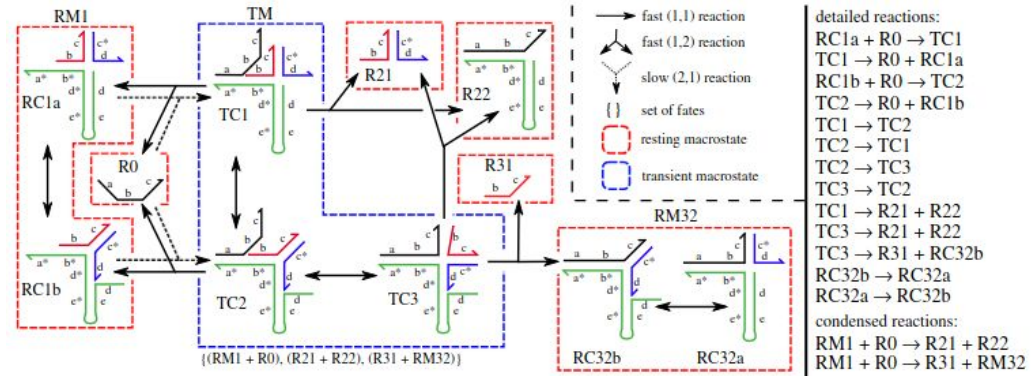
Energy Landscape



Badelt, Flamm, Hofacker (2016)



Reaction enumeration for DNA computation: peppercornenumerator

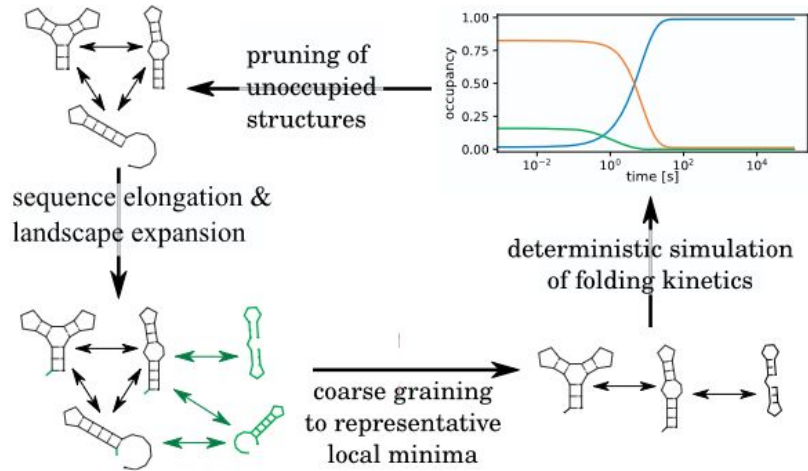


Badelt, Grun, Sarma, Wolfe, Winfree (2020)



... and in more recent history

co-transcriptional folding: **DrTransformer**

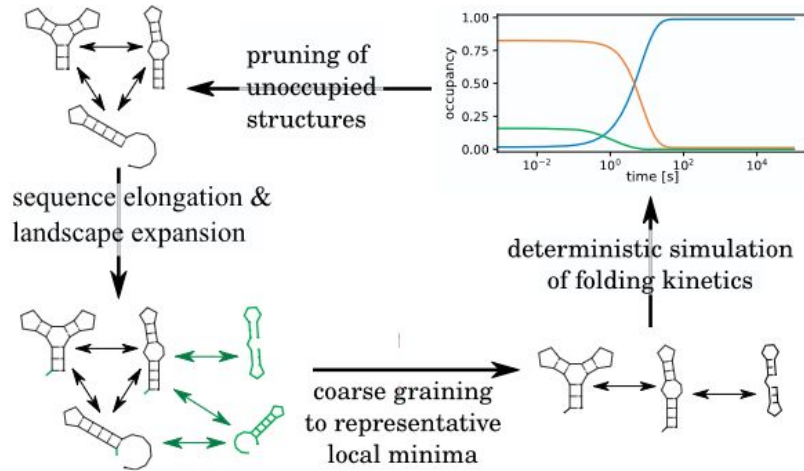


The main problem for DrTransformer:

Badelt, Lorenz, Hofacker (2022)

... and in more recent history

co-transcriptional folding: **DrTransformer**



Badelt, Lorenz, Hofacker (2022)

The main problem for DrTransformer:

- It has to express dynamics in an exponentially growing secondary structure landscape.

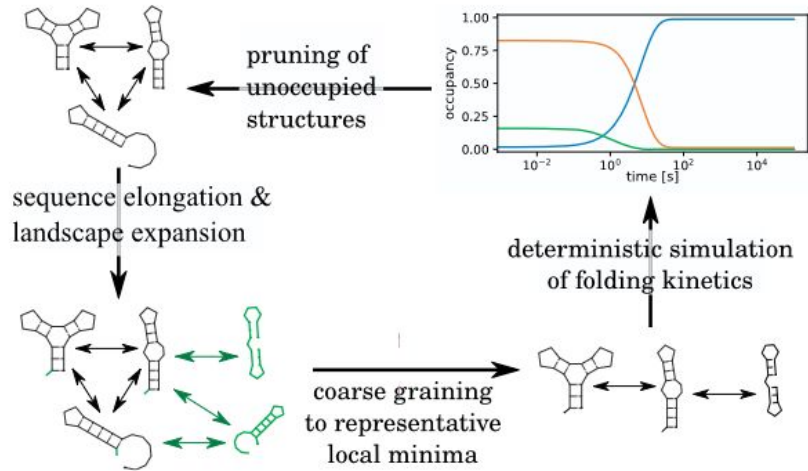
This problem is very common:

- Kinfold
- Kinefold
- barriers / treekin
- RNAexplorer
- RNALocmin
- ...

It is time to change the perspective!

... and in more recent history

co-transcriptional folding: **DrTransformer**



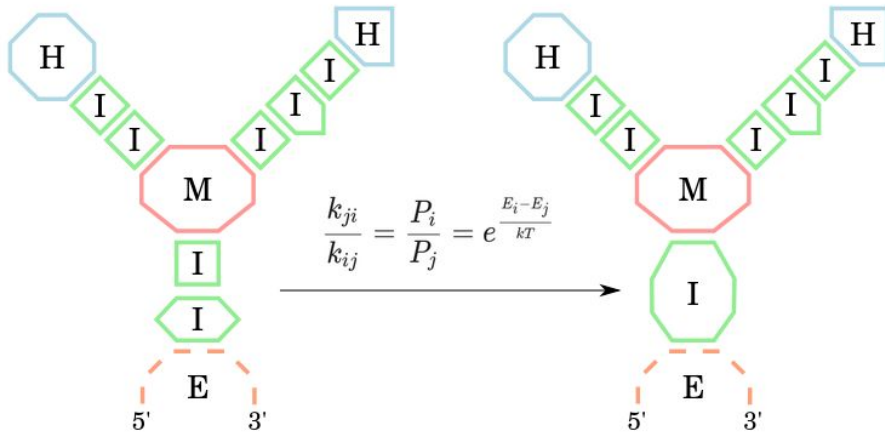
Today

- 1) Revert to the most detailed model to analyze secondary structure folding kinetics
- 2) Introduce parameters beyond k_0 toward a better biophysical model of RNA/DNA folding kinetics
- 3) Co-transcriptional simulations of 3kb transcripts
- 4) Co-transcriptional simulations in RNA design objective functions.

Badelt, Lorenz, Hofacker (2022)

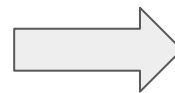
Basics of stochastic RNA folding kinetics

Choose a **move set** and a **rate model** that satisfies detailed balance, e.g.:

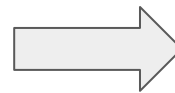


Simulate stochastic folding trajectories as given by the master equation

$$\frac{dP_i(t)}{dt} = \sum_{i \neq j} (P_j(t)k_{ji} - P_i(t)k_{ij})$$



yields the correct equilibrium distribution



whether kinetics is correct depends on the rate model!

ff-trajectory: single stochastic simulations

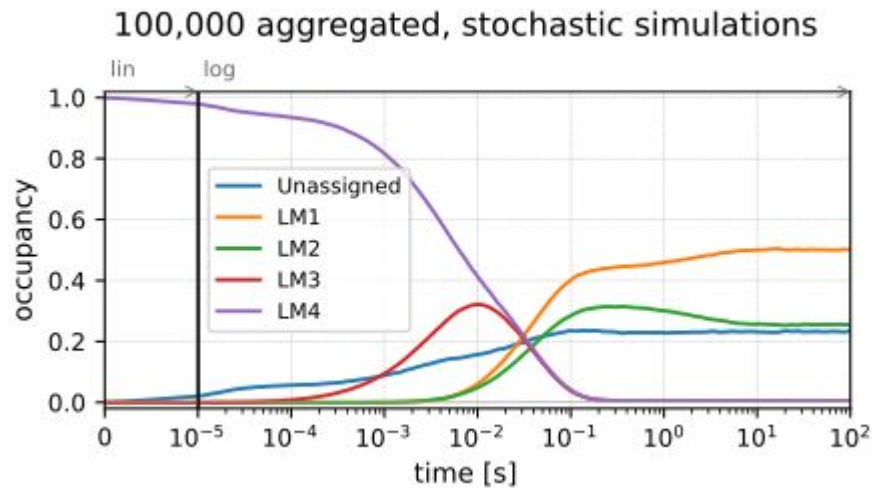
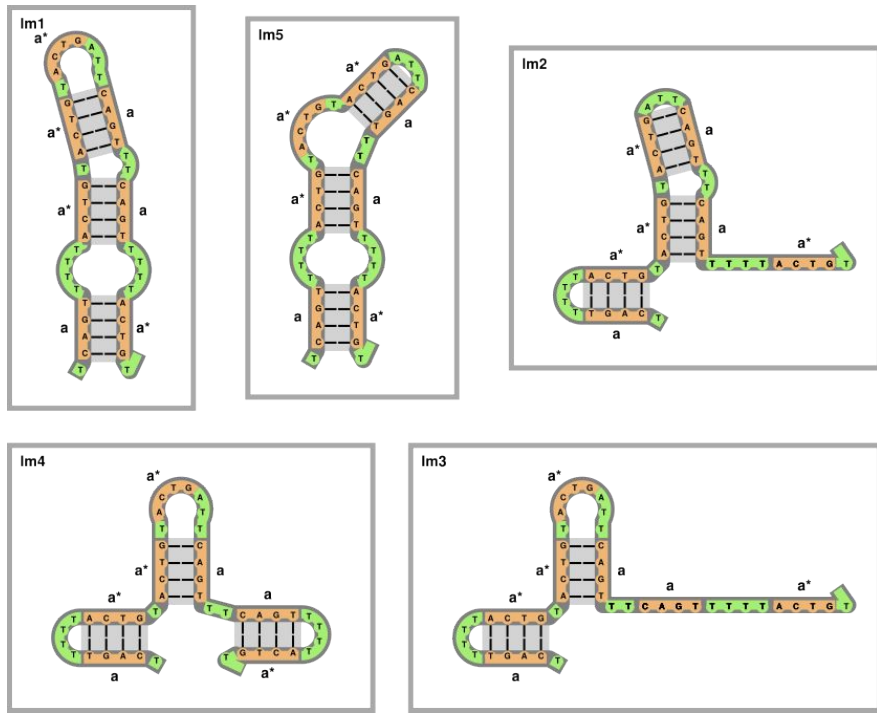
GCGUUUCCAGGGUUUAGACGGACGGGUGUGACUCGCCCAGCCCCGACCUC	energy	arrival-time	waiting-time	mean-waiting
.....	0.00	0.00000000e0	5.48838308e-1	1.15990269e0
.....(.....)	2.10	5.48838308e-1	7.42186429e-1	6.00461698e-1
.....(.....).....(.....)	4.50	1.29102474e0	6.62519714e-1	3.57034800e-1
.....(.....)	2.40	1.95354445e0	8.70308936e-3	5.15932279e-1
.....	0.00	1.96224754e0	2.24827455e0	1.15990269e0
.....(.....)	1.90	4.21052209e0	1.08942365e0	4.14255646e-1
.....((.....))	0.10	5.29994574e0	3.93454939e-1	5.63963354e-1
.....((.....)).	-0.20	5.69340068e0	1.06769149e-1	4.52730587e-1
.....((.....)).	1.80	5.80016983e0	4.33243095e-1	3.16296938e-1
.....(((.....)))	-2.40	6.23341293e0	1.28813977e0	8.21431625e-1
.....((((.....))))	-1.60	7.52155270e0	2.45415452e-1	2.40731917e-1
.....((((.....)))	-1.50	7.76696815e0	6.94202383e-2	4.10851300e-1
.....((((.....))))	-5.60	7.83638839e0	3.47571920e0	3.26611240e0
.....((((.....)))	-2.90	1.13121076e1	1.55874453e0	4.13337266e-1
.....((((.....))))	-5.60	1.28708521e1	2.36328447e0	3.26611240e0
.....((((.....))))(.....)	-2.10	1.52341366e1	7.41356896e-3	4.84982503e-1
.....((((.....))))((.....))	-4.70	1.52415502e1	2.00829892e1	1.23961105e1
..(.....((((.....))))((.....))	-0.50	3.53245394e1	1.26546682e0	7.99897165e-1
.....((((.....))))((.....))	-4.70	3.65900062e1	4.58067121e1	1.23961105e1

ff-trajectory: single stochastic simulations

GCGUUUCCAGGGUUUAGACGGACGGGUGUGACUCGCCCAGCCCCGACCUC	energy	arrival-time	waiting-time	mean-waiting
.....	0.00	0.00000000e0	5.48838308e-1	1.15990269e0
.....(.....)	2.10	5.48838308e-1	7.42186429e-1	6.00461698e-1
.....(.....).....(.....)	4.50	1.29102474e0	6.62519714e-1	3.57034800e-1
.....(.....)	2.40	1.95354445e0	8.70308936e-3	5.15932279e-1
.....	0.00	1.96224754e0	2.24827455e0	1.15990269e0
.....(.....)	1.90	4.21052209e0	1.08942365e0	4.14255646e-1
.....((.....))	0.10	5.29994574e0	3.93454939e-1	5.63963354e-1
.....((.....)).	-0.20	5.69340068e0	1.06769149e-1	4.52730587e-1
.....((.....)).	1.80	5.80016983e0	4.33243095e-1	3.16296938e-1
.....(((.....)))	-2.40	6.23341293e0	1.28813977e0	8.21431625e-1
.....(((.....)))	-1.60	7.52155270e0	2.45415452e-1	2.40731917e-1
.....(((.....)))	-1.50	7.76696815e0	6.94202383e-2	4.10851300e-1
.....((((.....))))	-5.60	7.83638839e0	3.47571920e0	3.26611240e0
.....((((.....))))	-2.90	1.13121076e1	1.55874453e0	4.13337266e-1
.....((((.....))))	-5.60	1.28708521e1	2.36328447e0	3.26611240e0
.....((((.....))))(.....)	-2.10	1.52341366e1	7.41356896e-3	4.84982503e-1
.....((((.....))))((.....))	-4.70	1.52415502e1	2.00829892e1	1.23961105e1
..(.....((((.....))))((.....))	-0.50	3.53245394e1	1.26546682e0	7.99897165e-1
.....((((.....))))((.....))	-4.70	3.65900062e1	4.58067121e1	1.23961105e1

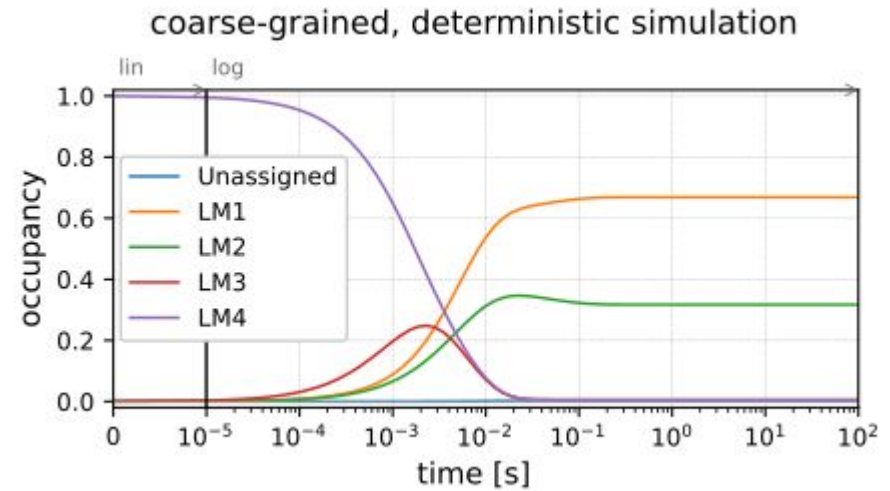
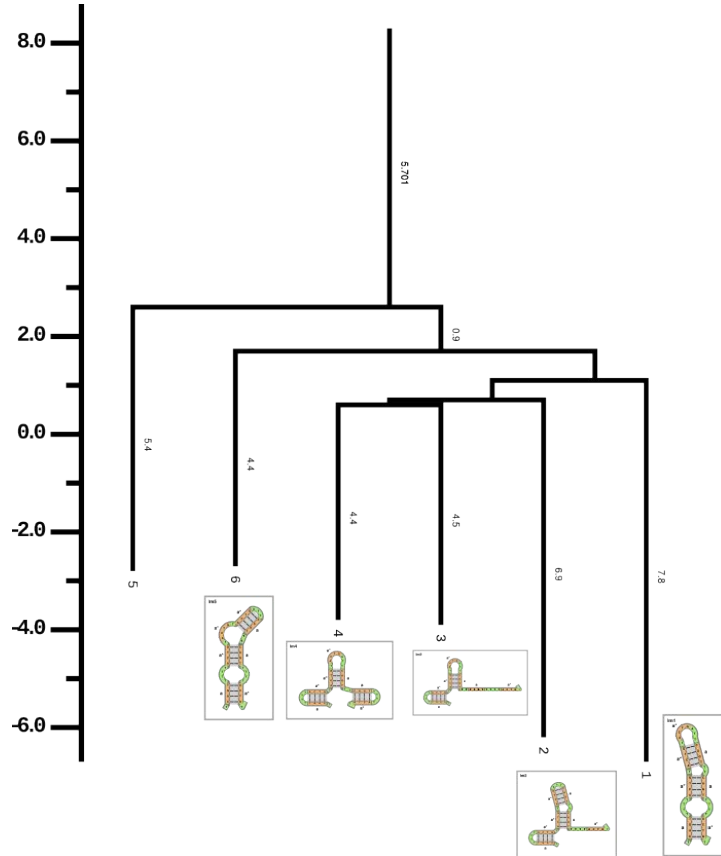
... but isn't that what Kinfold (Flamm et al. 2001) does?

ff-timecourse: use your own coarse graining!



Dynamics of LM structures including local neighbors within 3 kcal/mol.

A comparison with a deterministic approximation



The barriers / treekin approach expresses dynamics between local minima.

Three-way branch migration may use shift moves

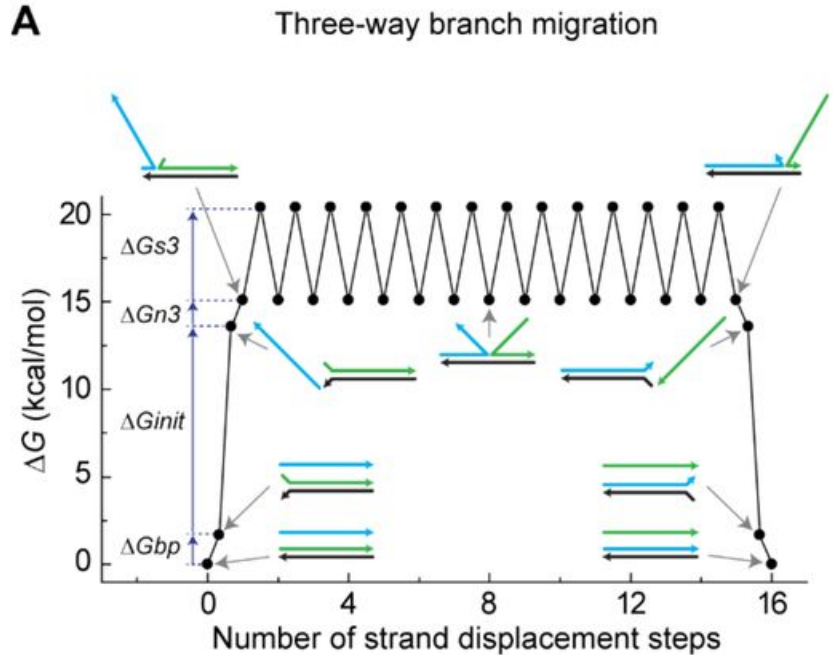
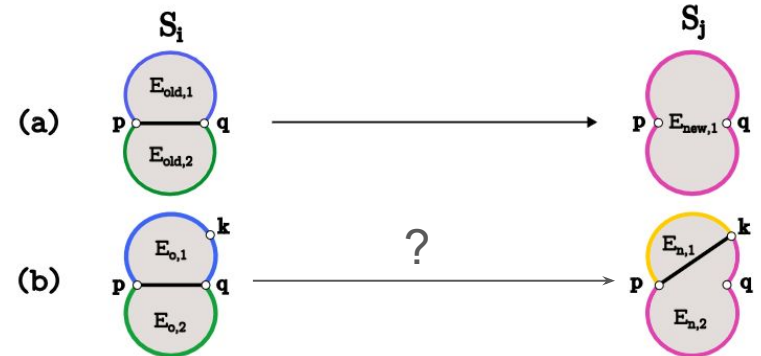


Figure from Kotani & Hughes (2017)

Two move types according to oxRNA Srinivas et al. (2013)

- "branch migration via sequential disruption and formation of bonds"
- "branch migration via invasion" (shift-moves!)



Three-way branch migration may use shift moves

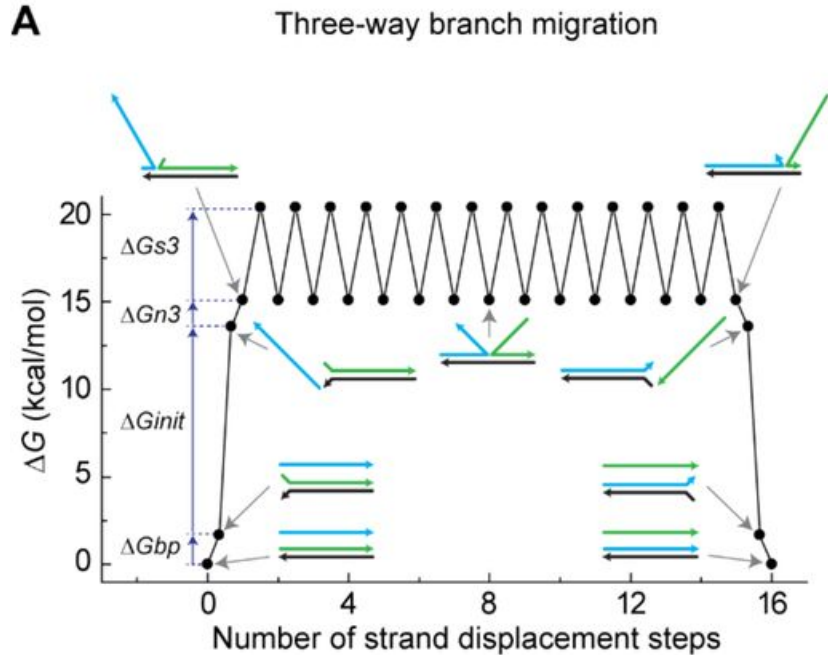
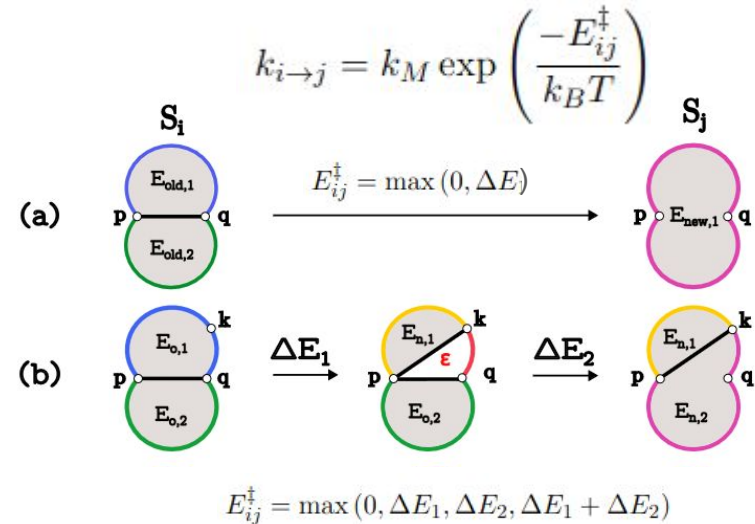


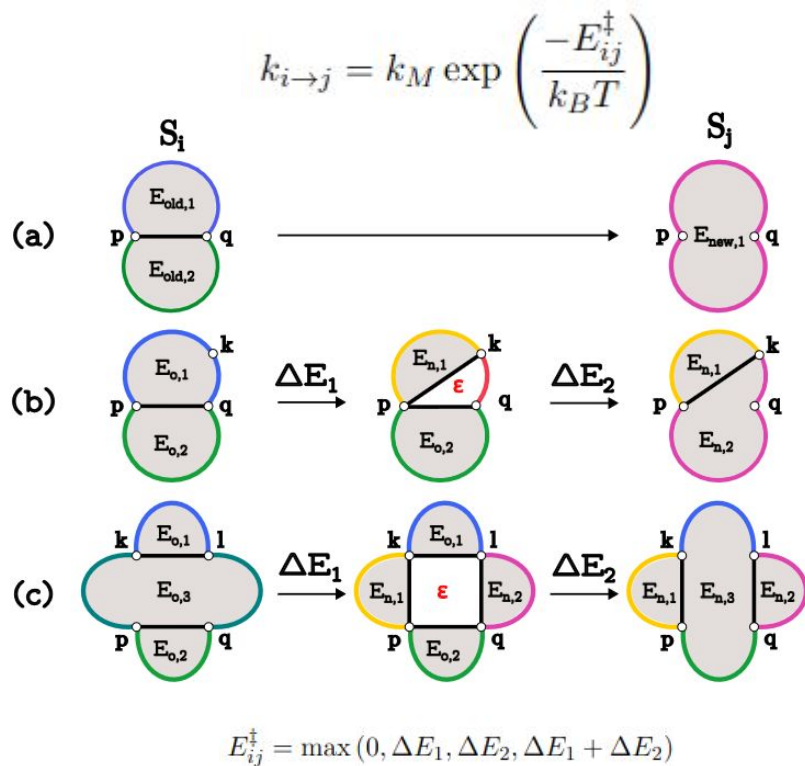
Figure from Kotani & Hughes (2017)

Two move types according to oxRNA Srinivas et al. (2013)

- "branch migration via sequential disruption and formation of bonds"
- "branch migration via invasion" (shift-moves!)

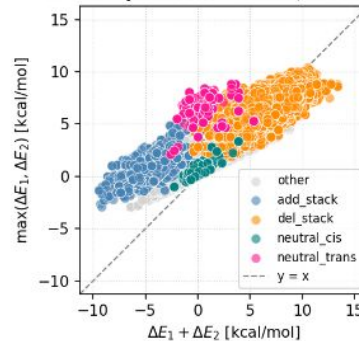


Shift-move parameterization matters!

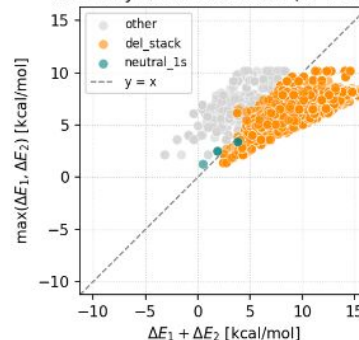


RNA design did1 ($\Delta E = 12$ kcal/mol)

three-way shift evaluation (n=630857)

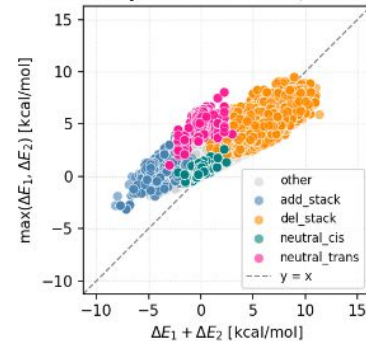


four-way shift evaluation (n=10156)

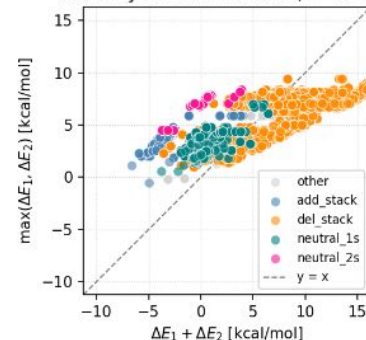


RNA design did2 ($\Delta E = 12$ kcal/mol)

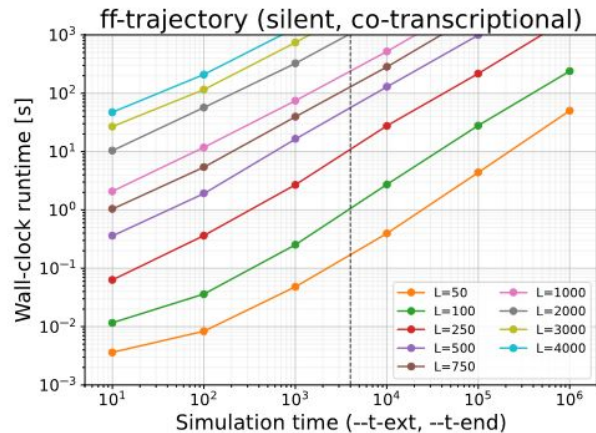
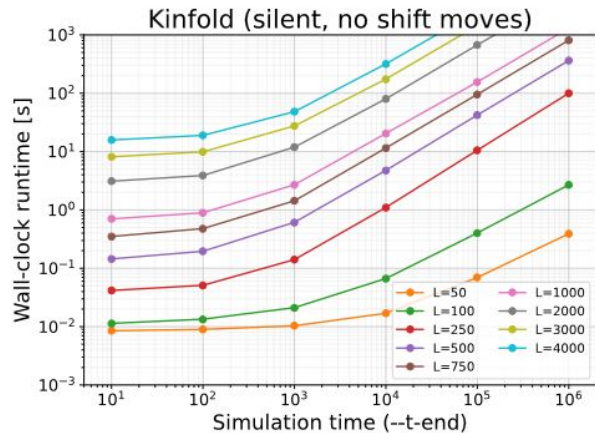
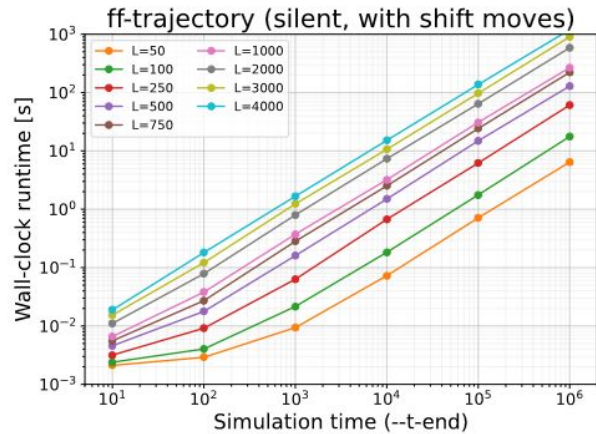
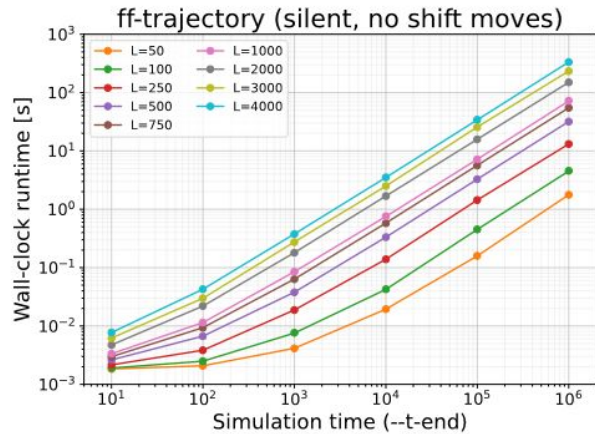
three-way shift evaluation (n=434458)



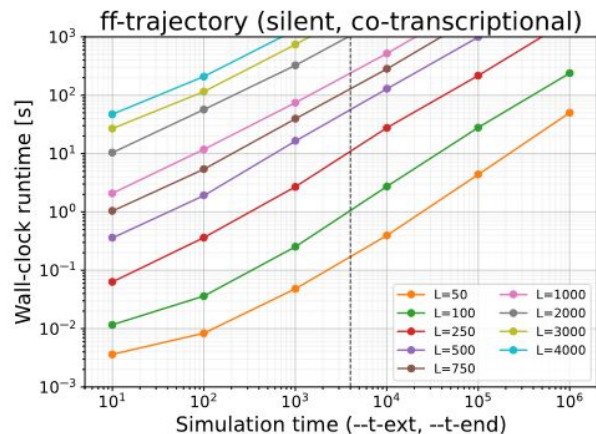
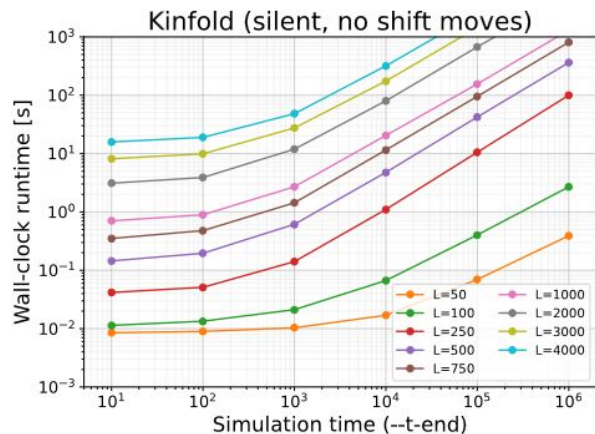
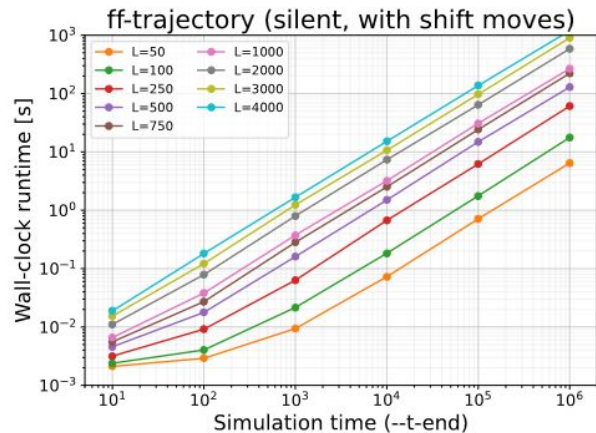
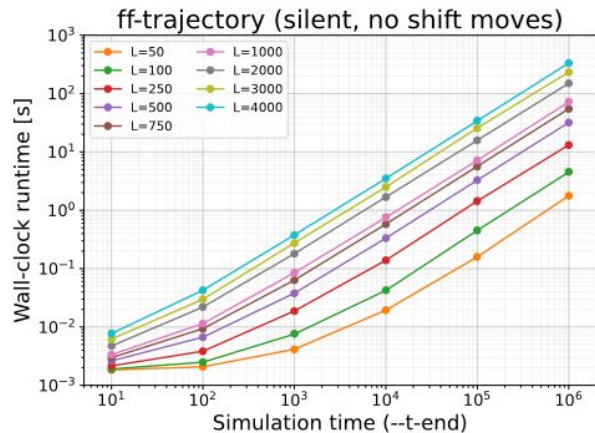
four-way shift evaluation (n=44065)



fuzzyfold is **very** fast



fuzzyfold is **very** fast

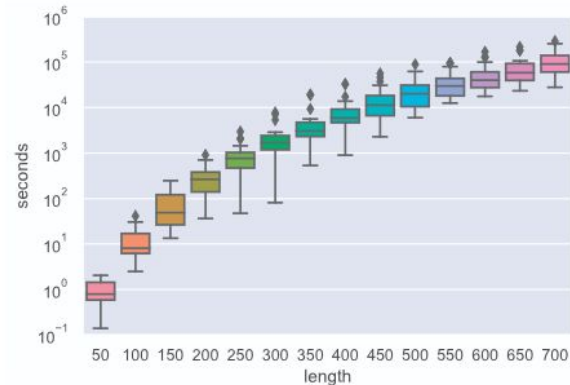


A co-transcriptional simulation of a 750 nt sequence takes:

- 1 min per trajectory

or

- 1 day for DrTransformer





Alain Laederach



Hannah Hartung

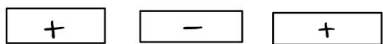
Are inverse ALU elements competing for structure?

~300nt long, non-autonomous retrotransposons

Approximately 10 percent of human genome

Sequence similarity between any two Alu elements > 80 %

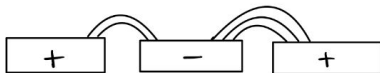
Inverted repeat Alu elements: high degree of complementarity



ALU triplets compete between different configurations.



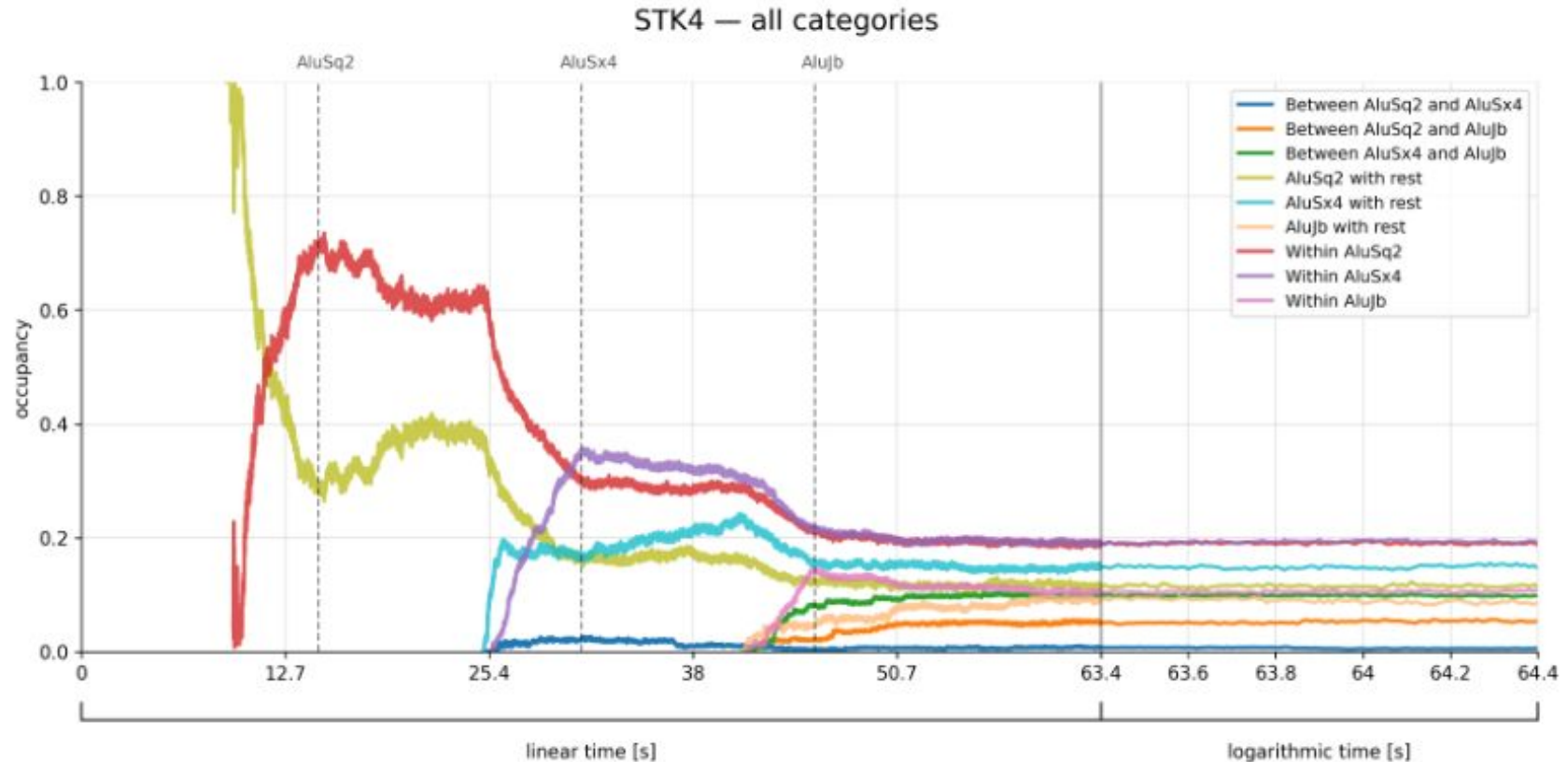
Impacts on splicing, A->I editing.



Experimental data on modifications exists, can we explain competitions With cotranscriptional folding.

Let's simulate examples (~3000 nt sequences) to see if ratio of base-pairs within ALUs and in-between ALUs depends on co-transcriptional dynamics.

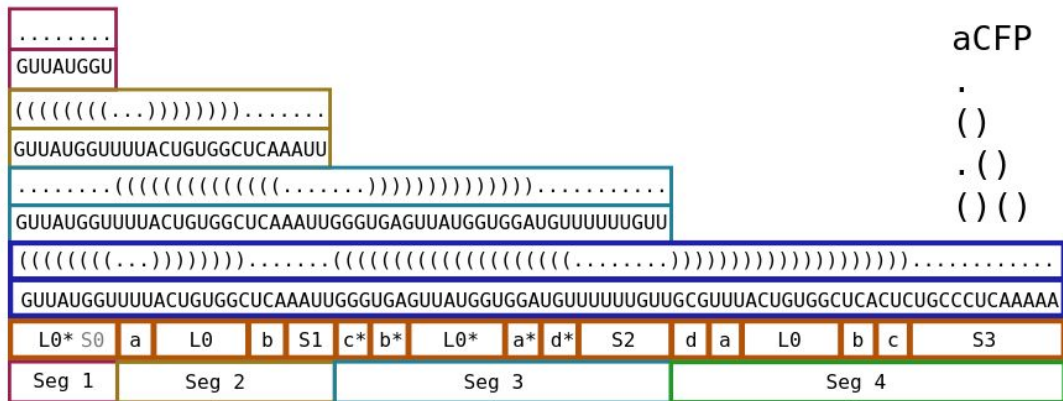
Fraction of base-pairs between ALUs



Next: see how Inosine parameters change the picture.

A portrait of a young man with short brown hair, wearing glasses and a grey crewneck sweater. He is looking directly at the camera against a plain, light-colored background.

Length 8: ...



Design of co-transcriptional folding pathways

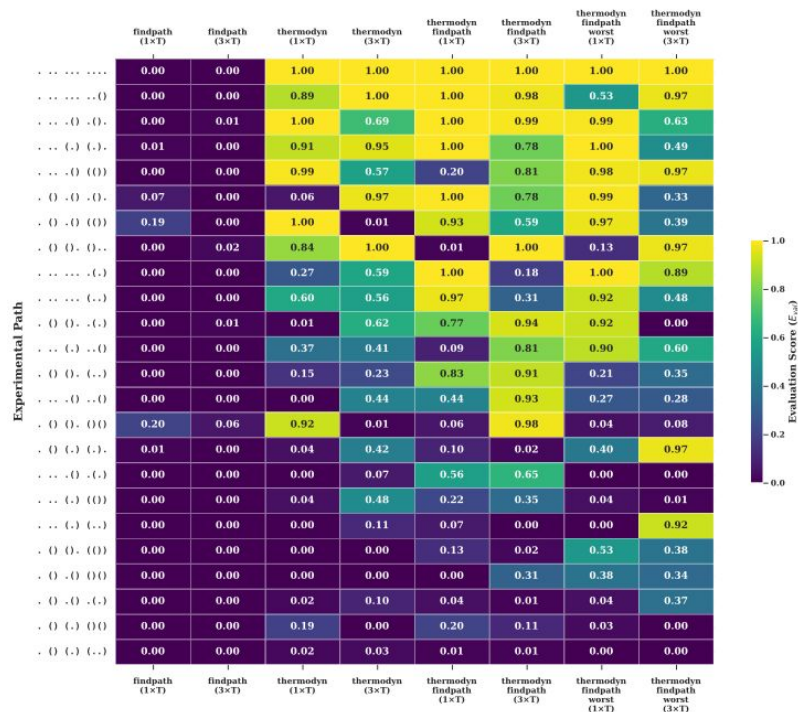


Daniel Guerguerian

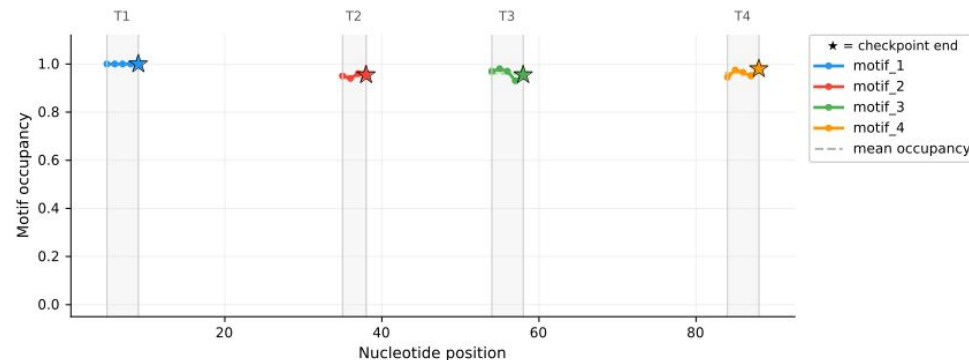
Thermodynamic design is extremely difficult!

Design using fuzzyfold trajectories.

Evaluation of Individual Paths: Flex Domain Length



. . . .() .()



Also under active development ...

Pseudoknot evaluation & integration



Jonas Jäger

Coaxial stacking



Anna Dohm

Thanks to

University of Vienna:

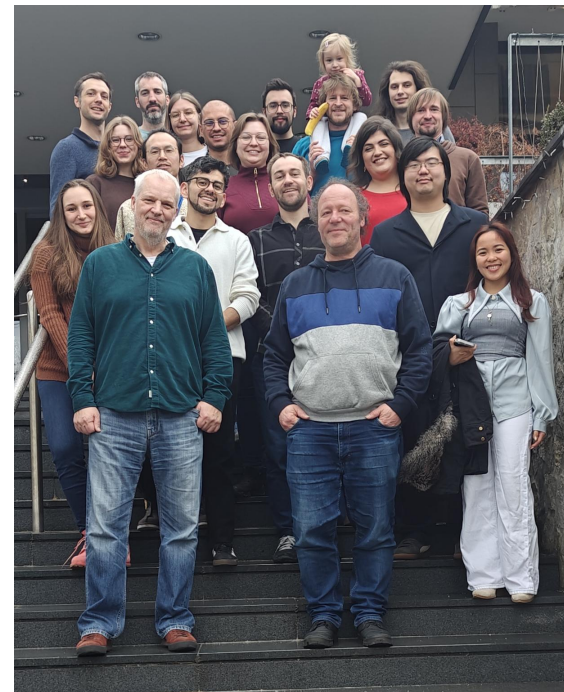
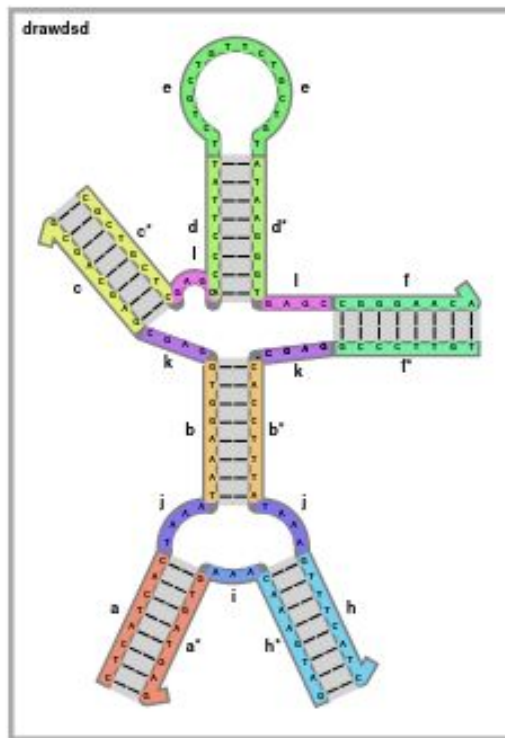
- Ivo L. Hofacker
- Hannah Hartung
- Daniel Guerguerian
- Jonas Jäger
- Anna Dohm
- Laurenz Miksch

UNC Chapel Hill:

- Alain Laederach
- Scott Allen

Caltech:

- Erik Winfree



tbi

Thanks to

University of Vienna:

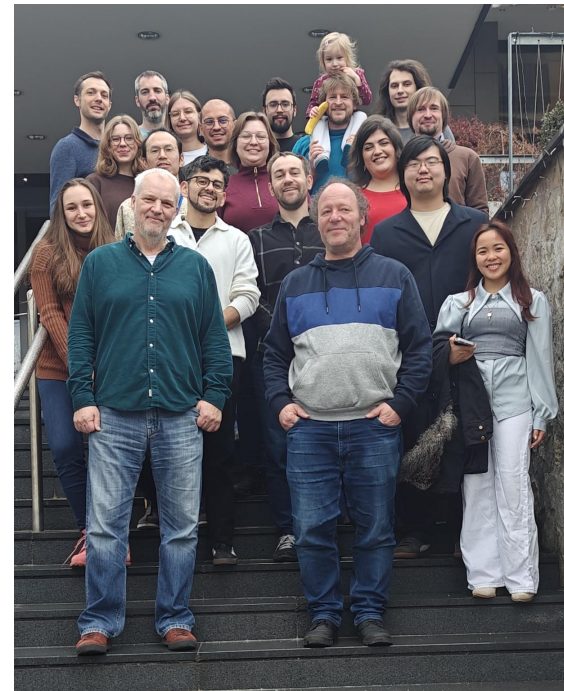
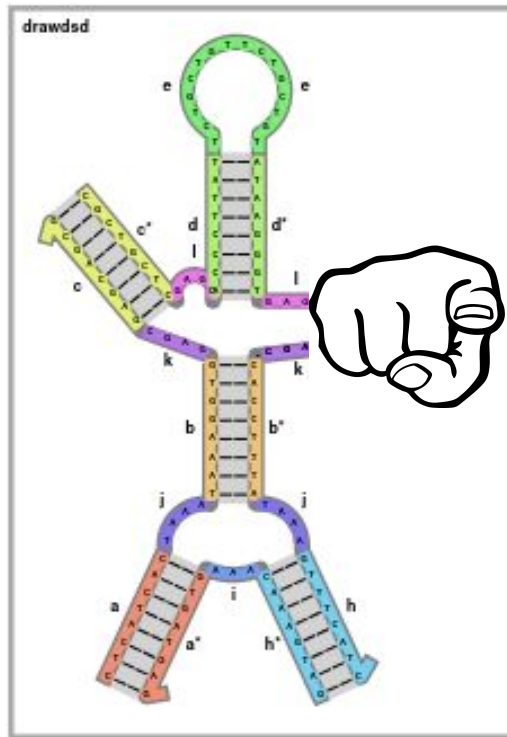
- Ivo L. Hofacker
- Hannah Hartung
- Daniel Guerguerian
- Jonas Jäger
- Anna Dohm
- Laurenz Miksch

UNC Chapel Hill:

- Alain Laederach
- Scott Allen

Caltech:

- Erik Winfree



tbi