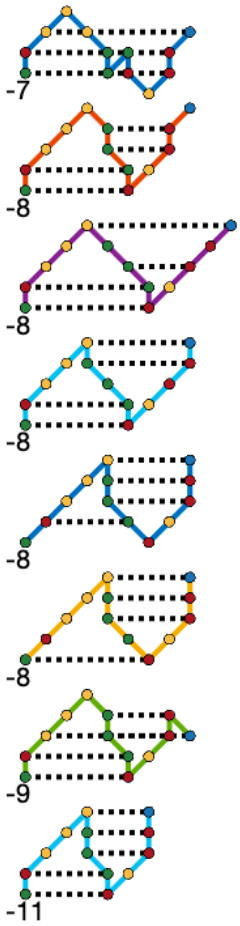
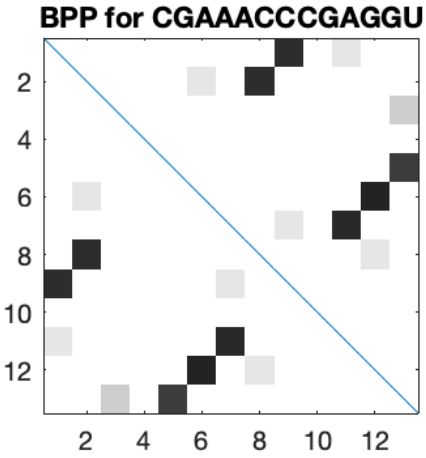
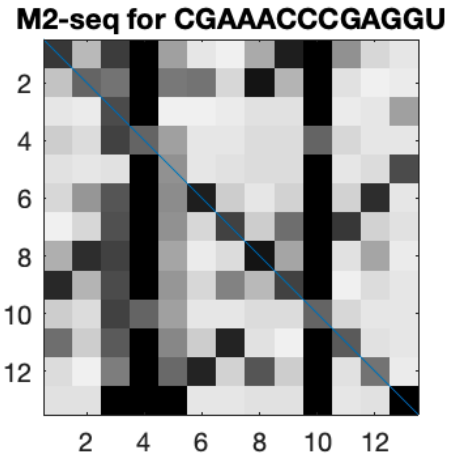
Revisiting ToyFold-1D
19 Jan, 2021

https://github.com/DasLab/ToyFold-1D

```
params = get_default_energy_parameters();
params.epsilon = -3;
params.delta = 2;
analyze_sequence('CGAAACCCGAGGU',params);
```





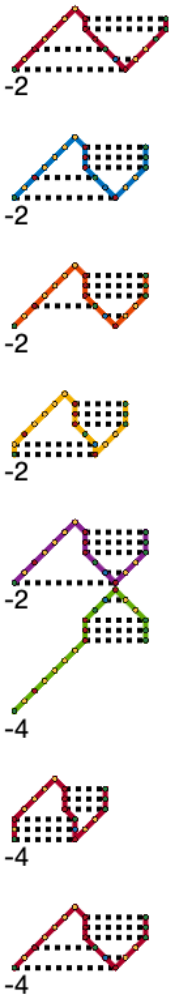


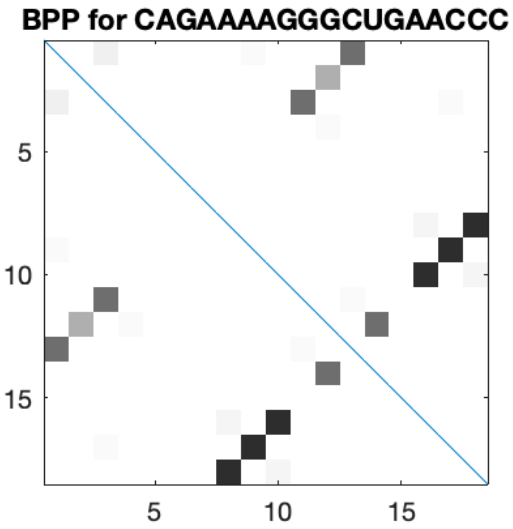
• Was useful last year in thinking through exactness of nearest neighbor rules.
• Raises prospect of generating 10^5 ‘gold standard’ MFE structures and

BPP which could be used for testing neural network architectures.
• including “tensor field networks” with signed features, and transformers to auto build coordinates of MFE.

I got rid of ‘direction’ vectors and constraint that base pairs must only form between residues moving in opposite direction. This was different from pencil-and-paper model and imposed weird constraints on structure (Forked that repo out to https://github.com/rhiju/ToyFold-1D-directed).

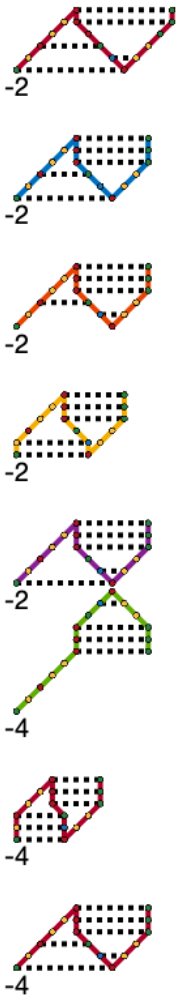
Test pseudo knot frequency
I actually coded this up in March 2020 — but didn’t preserve notes.
Now running again in new model

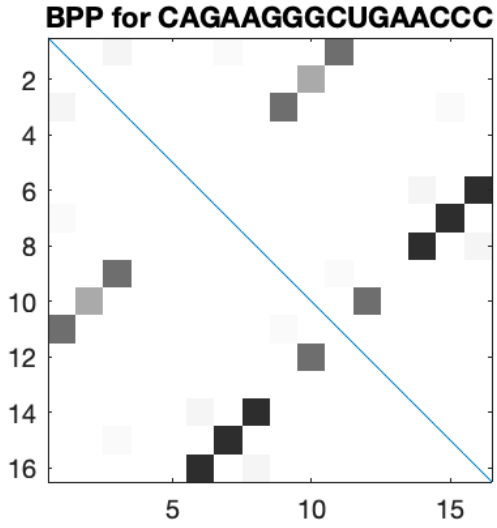




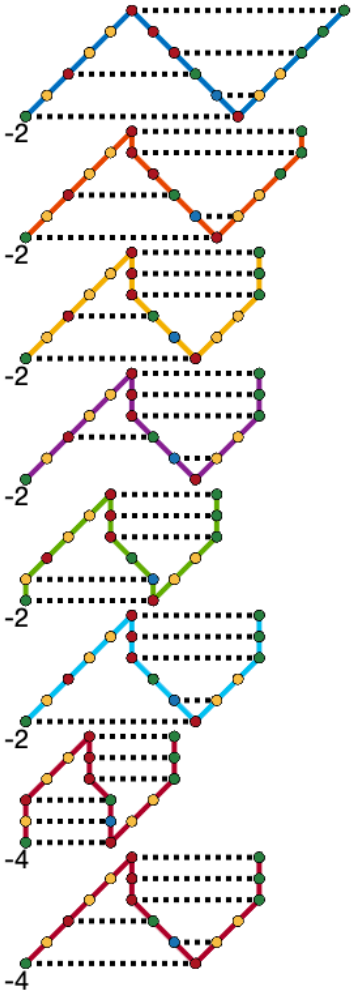
Took a long time — only 18 nts, but took:

Elapsed time is 79.262615 seconds.
3060413 conformations!
Also note that there's degeneracy lowest energy conformations.
Should be able to reduce length of first loop — no longer have that funny directionality constraint:





Filter for pseudo knot conformations only — yup that still works:

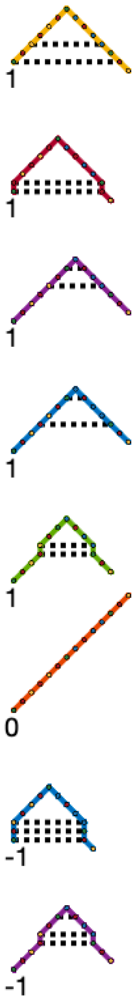


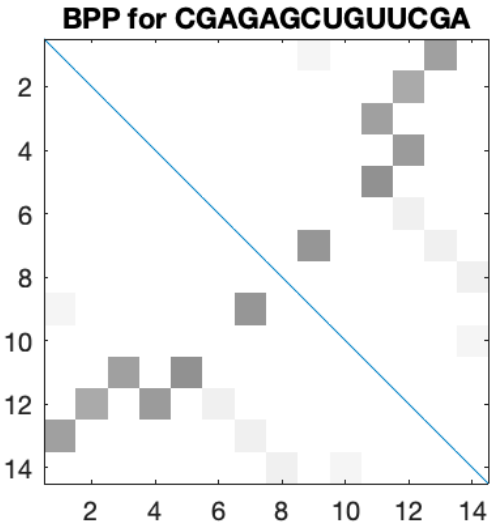
Get data for 14-nt random sequences:

params.epsilon = -2; params.delta = 5; % try to get more cooperativity

14-nt

Yea, not getting too many pseudo knots. This is a good starting point for generating a lot of train/test data?





Each sequence is taking 1-3 seconds.

4^14 = 268,435,456

If I can get under 1 sec, would take ~30 hours to generate 10^5 training data sets.

Profile: analyze_sequence('AGCGGACAGUCUGA',params,0);
Figured out some computations that weren't necessary based on profiling — got ~2x speedup to ensure ~1 sec time.