

# Kinwalker parameter space analysis

Mario Koestl

Institute for Theoretical Chemistry  
University of Vienna

June 16, 2016

# Overview

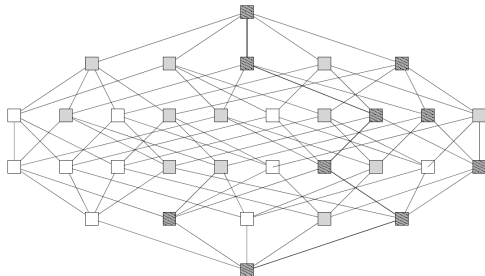
- ▶ co-transcriptional folding = RNA folding at the growing 5' end
- ▶ co-transcriptional pathway prediction difficult
- ▶ Kinwalker algorithm: heuristic approach for RNA folding kinetic prediction

# Kinwalker

- ▶ predicts secondary structure trajectories up to 1500 nt
- ▶ uses structures backtracked from  $C_{i,j}$  matrix after each base adding step.
- ▶ 2 direct pathfinding algorithm : "Morgan Higgs" and "findpath"

# Kinwalker parameters used

- ▶ transcription rate
- ▶ barrier heuristic
- ▶ dangle model
- ▶ max keep



## data used

- ▶ 200 different parameter combinations for 3 RNA families
- ▶ SRP, TRP and RNaseP
- ▶ alignment for each family from Rfam database, according to Meyer et al.
- ▶ reference structures taken from literature, reference sequences downloaded from NCBI

# Scoring functions

- ▶ comparing Kinwalker output data with different parameters
- ▶ Structure Conservation Index and Sequence Similarity
- ▶ Base-pair diversity, ensemble diversity, ensemble distance

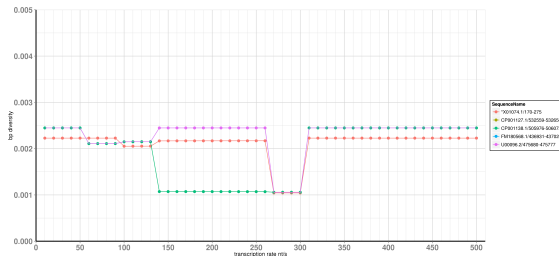
# ensemble distance

- ▶ expected distance between folding trajectory and a reference structure

$$\text{ensemble distance}(s, \text{ref}) = \sum_{i,j \in \text{ref}} (1 - p(i,j)) + \sum_{i,j \notin \text{ref}} p(i,j), \quad (1)$$

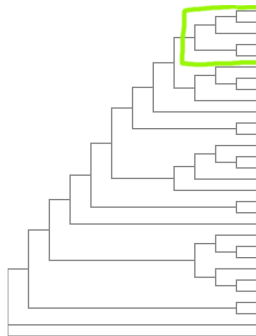
- ▶  $p(i,j) = \frac{\Delta \text{Time}}{T_{\text{total}}}.$

# Results



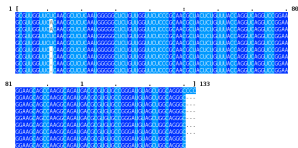
sequence	trap at 260 nt/s	trap at 300 nt/s	trap at 310 nt/s
X01074.1.170-275	yes	no	yes
U00096.2.475680-475777	yes	no	yes
FM180568.1.436931-437028	yes	no	yes
CP001138.1.505976-506073	yes	no	yes
CP001127.1.532559-532656	no	no	yes





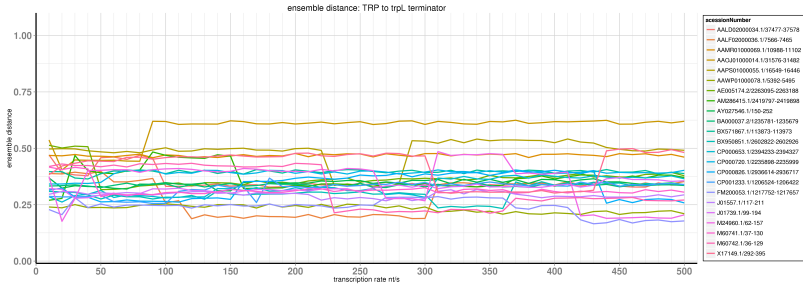
X01074\_1/170-275 0  
 U00096\_2/475680-475777 0  
 FM180568\_1/436931-437028 0  
 CP001127\_1/532559-532656 0  
 CP001138\_1/565976-566073 0  
 AE008952\_1/117912-117929 0.00398  
 CP001048\_1/1157694-1157791 0  
 CP000720\_1/3460787-3460890 0  
 AL590842\_1/3489558-3489661 0.01895  
 CP000671\_1/1242908-1242909 0.21571  
 AE016795\_3/3210836-3219740 0.10027  
 CP000020\_2/1087907-1087810 0.11067  
 CP000076\_1/2115321-2115418 0.03583  
 CP000075\_1/2101610-2101707 0  
 CP000058\_1/2086711-2086806 0  
 CP000712\_1/1786696-1786793 0.09003  
 X14404\_1/127-224 0.07905  
 CP000013\_1/484623-484720 0.31653  
 BX071662\_1/289282-289380 0.32463  
 AL157959\_1/1147858-1147765 0.21815  
 CP000086\_1/2494952-2495047 0.00913  
 CP000010\_1/1421860-1421865 0  
 CP000570\_1/2220068-2220093 0  
 CP000090\_1/2325066-2325165 0.08217  
 CP001068\_1/1100816-1100717 0.04415  
 CU833749\_1/2046219-2046315 0.04179  
 AC1501000002\_1/2525660-2525659 0.08626  
 AE016825\_1/2560853-2560754 0.09556  
 CP000116\_1/544561-544656 0.1373  
 CP000450\_1/597363-597461 0.12052

1 X01074\_1/170275 100.0%  
 2 CP001127\_1/5325 99.2%  
 3 CP001138\_1/5259 99.2%  
 4 U00096\_2/475680 100.0%  
 5 FM180568\_1/4369 100.0%  
 consensus/100%  
 consensus/90%  
 consensus/80%  
 consensus/70%  
 1 X01074\_1/170275 100.0%  
 2 CP001127\_1/5325 99.2%  
 3 CP001138\_1/5259 99.2%  
 4 U00096\_2/475680 100.0%  
 5 FM180568\_1/4369 100.0%  
 consensus/100%  
 consensus/90%  
 consensus/80%  
 consensus/70%



# TRP family

## ► Ensemble Distance for trp sequences to terminator structure



# conclusion

blablabla, auch was das es halt icht geht