Kinwalker parameter space analysis
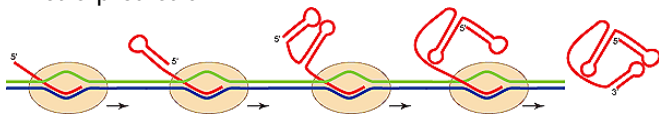
Mario Köstl

Institute for Theoretical Chemistry
University of Vienna

February 16, 2016

Introduction/Motivation

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

- Questions: Existence of evolutionary conserved folding pathways? Can Kinwalker predict them?

- predicts secondary structure trajectories up to 1500 nt
- find a way from the $i - j$ to the $i - (j+1)$ structure
- uses direct path-finding heuristics

```
GCGGUGGUCUCAACGCUCUCAUGGGGGCUCUGUUGGGUGCUCCCCAACGCUACUCUGUUUACCAGGUCAGGUCCGGAAGGAAGCAGCCAAGGCAGAUGACGCGUGUGCCGGGAUGUAGCUGGCAGGGCCCC
((((((.....)))))).......((((((((((((((((((((((((...(((...(((....)))....)))....))))))))).).))))).)).)))))).).)))))))))))))) -68
TRAJECTORY
..((((.....))))) -1.3 0.789999 2.7 7.83608 15
.((((.....))))) -3.6 0.75 0 7.83608 16
(((((.....)))))) -6.8 0.8 0 7.83608 17
((((((.....)))))) -0.5 1.40081 2.7 7.83608 29
((((((.....)))))).((((((.....)))))) -10.1 1.80013 4.3 7.83608 37
((((((.....)))))).......((((((.....)))))) -12.5 2.12480 7.4 7.83608 43
((((((.....)))))).......((((((...))...)))))) -14.1 2.28323 6.2 7.83608 45
((((((.....)))))).......((((.((...))...)))))) -16.2 3.35002 3.2 7.83608 68
((((((.....)))))).......((((.((...))...)))))...........((((...))))) -16.9 3.65019 4.5 7.83608 74
((((((.....)))))).......((((.((...))...)))))...........((((.(((.....))))).)))) -17.8 3.65037 4.5 7.83384 74
((((((.....)))))).......((((.((...))...)))))...(((.(((((.....)))))).)))) -17.9 3.65038 0.799999 7.83159 74
((((((.....)))))).......((((.((...))...)))))...(((.(((((.....)))))).)))) -19.7 3.90195 5.9 7.83608 79
((((((.....)))))).......((((.((...))...)))))....(((((.((((....))).)))).)))).)).))) -22.3 4.25081 2.5 7.83608 86
((((((.....)))))).......((((.((...))...)))))....(((((.((((...)))).)).))).)).))) -23.8 4.7 2.2 7.83608 95
((((((.....)))))).......((((.((...))...)))))....(((((.((((...)))).)).))).)).))) -24.2 4.70802 3 7.83604 95
((((((.....)))))).......((((.((...))...)))))....((((((.((((....)))).)).))).((..))).))) -24.3 5.25003 3.4 7.83608 106
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))).))))).))) -27.0 5.45831 4.8 7.83608 110
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))).)))).))) -31.3 5.5 0 7.83608 111
((((((.....)))))).......((((.((...))...)))))....((((((((.((((....)))).)).))))).))).))) -33.5 6.00003 3.4 7.83608 121
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))))).))).)).))) -34.9 6.35005 5.4 7.83608 128
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))))).))).)).))) -35.7 6.36808 7.2 7.8259 128
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))))).))).)).))) -37 6.36808 1 7.56811 128
((((((.....)))))).......((((.((...))...)))))....(((((((.((((....)))).)).))))).))).)).))) -38.8 6.40003 3.4 7.83608 129
((((((.....)))))).......((((.((...))...)))))....((((((((.((((....)))).)).))))).))).)).))) -48.6 33.9127 11.6 12 133
((((((.....)))))).......((((((((.((...))...)))))...(((((((.((((....)))).)).))))).))).)).))) -41 33.9128 4.3 12 133
((((((.....)))))).......(((((((((.((...))...)))))...(((((((.((((....)))).)).))))).))).)).))) -43.1 72.8913 11.8 12 133
((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -45.9 2.77738e+15 30.06 31 133
((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -47.6 2.77738e+15 2.7 31 133
.((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -47.9 2.77738e+15 9 31 133
.((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -48.2 2.77738e+15 5.9 31 133
.((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -54.3 2.77738e+15 5.2 31 133
((((((.....)))))).......(((((((((((((((((((((((...(((((((.((((....)))).)).))))).))).)).)))))))))) -68 2.77738e+15 6.2 31 133
Kinwalker run time:33.09 seconds
```

Kinwalker parameters used

- ▶ transcription rate
- ▶ barrier heuristic : Morgan Higgs, findpath
- ▶ dangle model
- ▶ max keep
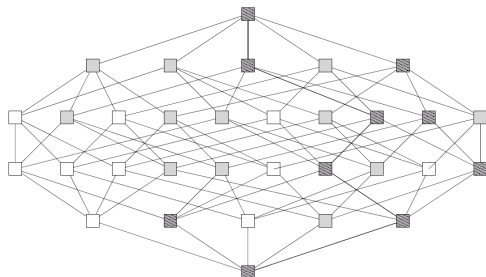
data acquisition

- ▶ 200 different parameter combinations for 3 RNA families
- ▶ SRP, TRP and RNAseP
- ▶ RFAM database for family alignments
- ▶ reference structures 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 of sequence s and a reference structure ref

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

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

Results SRP
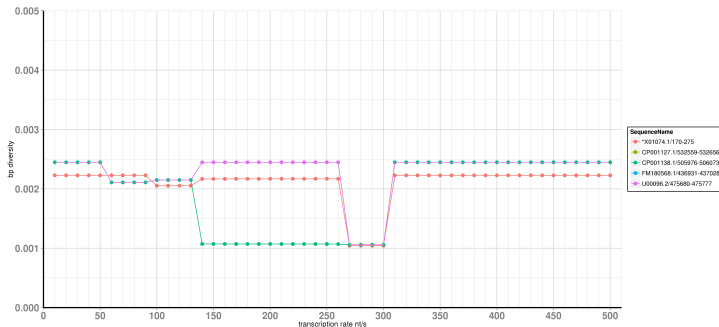
$$bpDiversity_{\Delta Time} = \frac{1}{maxBP} \cdot \sum_{(i,j)} \begin{cases} \frac{\Delta Time}{T_{total}} & \text{if (i,j) has been formed} \\ 0 & \text{if (i,j) has not been formed} \end{cases}$$
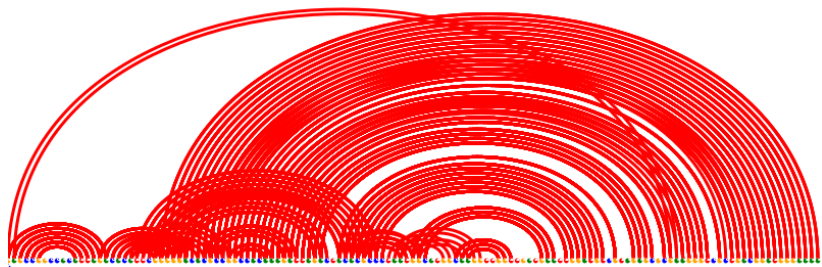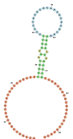
- clustered into subtree
- no compensatory mutations
- no evidence for co-transcriptional folding conservation
- evolutionary sequence conservation
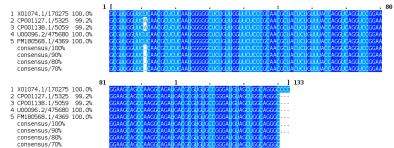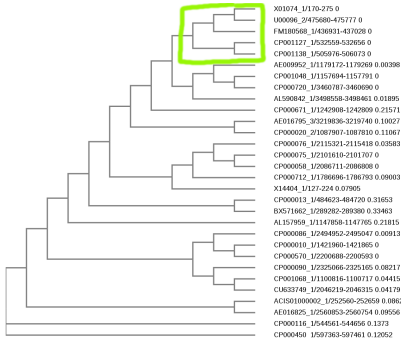
Results TRP

- ▶ Ensemble Distance for trp sequences to terminator hairpin structure

ensemble distance: TRP to trpL terminator
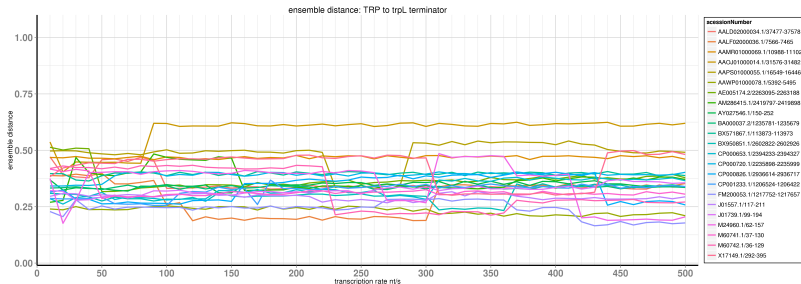
High level of tryptophan

complete leader peptide — ribosome — Stem loop: transcription termination signal — RNA pol

5' — mRNA — 1 — 2 — 3 — 4 — UUUU 3'

Trp codons

DNA — trpL

Low level of tryptophan

incomplete leader peptide — ribosome stalls — Alternate stem loop: transcription continues — RNA pol

5' — 1 — 2 — 3 — 4 — trp regulated genes

Trp codons

DNA — trpL
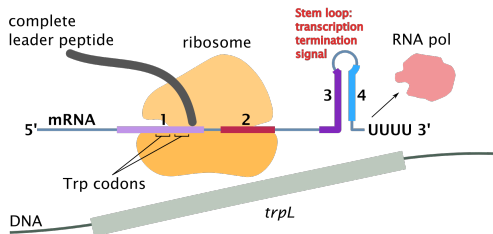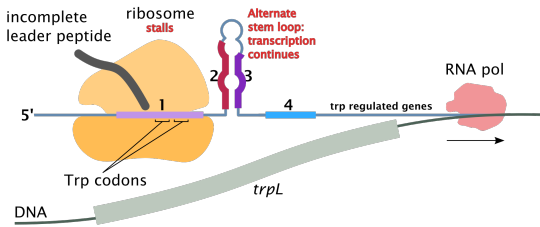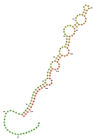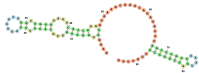
Conclusion

No evidence of co-transcriptional folding conservation was found

1. There really is no co-transcriptional folding conservation
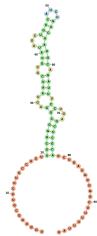2. There is co-transcriptional folding conservation, but I cannot see it

Thank you for your attention!

X01074.1 functional structure	X01074.1 transient structure
AE005174.2 alternative 1	AE005174.2 alternative 2
CP001509.3 functional	

trpL terminator
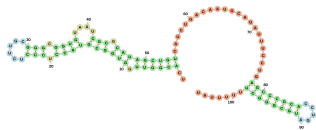
trpL anti-terminator

Bacteria
  Proteobacteria
    Betaproteobacteria
      Burkholderiaceae
        Cupriavidus
          Cupriavidus taiwanensis
          Ralstonia eutropha JMP134
        Ralstonia pickettii 12J
        pseudomallei group
          Burkholderia pseudomallei 668
          Burkholderia thailandensis E264
          Burkholderia mallei ATCC 23344
      Nitrosomonas eutropha C91
      Thiobacillus denitrificans ATCC 25259
      Neisseriales
        Chromobacteriaceae
          Pseudogulbenkiania ferrooxidans 2002
          Chromobacterium violaceum ATCC 12472
        Neisseria meningitidis Z2491 (Neisseria meningitidis serogroup A strain Z2491)
    Gammaproteobacteria
      Enterobacteriaceae
        Escherichia coli
          Escherichia coli O127:H6 (Escherichia coli 0127:H6)
          Escherichia coli str. K-12 substr. MG1655
        Yersinia pseudotuberculosis complex
          Yersinia pseudotuberculosis
            Yersinia pseudotuberculosis PB1/+
            Yersinia pseudotuberculosis IP 31758
          Yersinia pestis
            Yersinia pestis CO92
            Yersinia pestis KIM10+
        Salmonella enterica subsp. enterica
          Salmonella enterica subsp. enterica serovar Agona str. SL483
          Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633
        Haemophilus influenzae PittEE