Sibiria

Ulrike Mückstein, Stephan Bernhart

Institute for Theoretical Chemistry and Molecular Structure Biology
University of Vienna

Bled, 2005

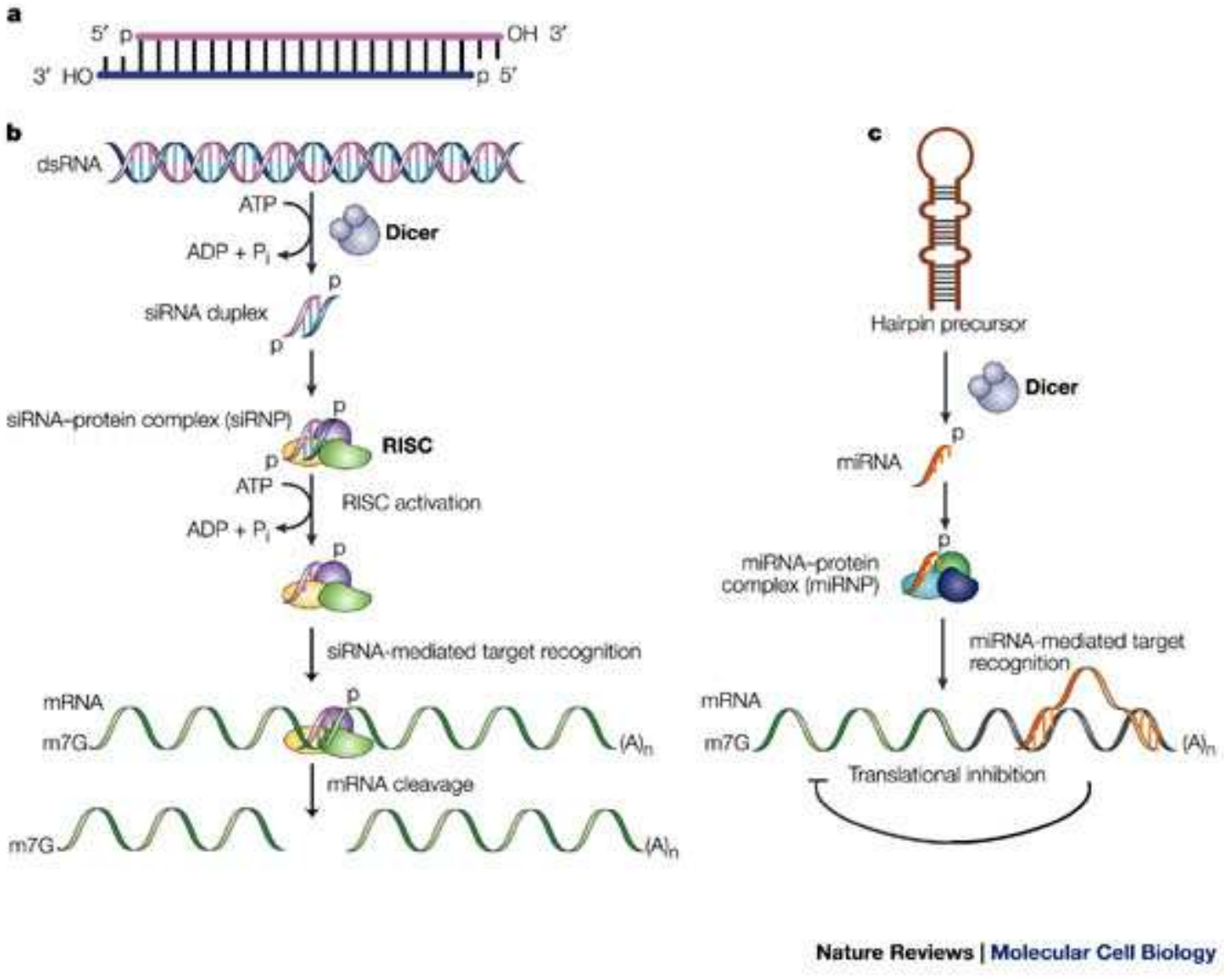
Outline

Outline

Why fold 2 RNA molecules

- Many RNA functions mediated by intermolecular RNA interactions
  - miRNA-siRNA pathway
  - RNA editing
- Design of custom-made RNA molecules
  - e.g. “openers” for binding sites

Intermolecular interactions of RNA molecules



2. Target accessibility and duplex thermodynamics

- Sfold by Ding & Lawrence
- Internal stability profile of small RNAs (5' and 3' end, average internal stability, . . .)
- probability profiling for prediction of unstructured regions in the target RNA:
  - generate a representative statistical sample of target RNA structures
  - calculate the probability that w consecutive positions, starting at position i, are within an unstructured region

2. Target accessibility and duplex thermodynamics

- Sfold by Ding & Lawrence
- Internal stability profile of small RNAs (5' and 3' end, average internal stability, . . .)
- probability profiling for prediction of unstructured regions in the target RNA:
  - generate a representative statistical sample of target RNA structures
  - calculate the probability that w consecutive positions, starting at position i, are within an unstructured region

2. Target accessibility and duplex thermodynamics

- Sfold by Ding & Lawrence
- Internal stability profile of small RNAs (5' and 3' end, average internal stability, . . .)
- probability profiling for prediction of unstructured regions in the target RNA:
  - generate a representative statistical sample of target RNA structures
  - calculate the probability that *w* consecutive positions, starting at position *i*, are within an unstructured region

2. Target accessibility and duplex thermodynamics

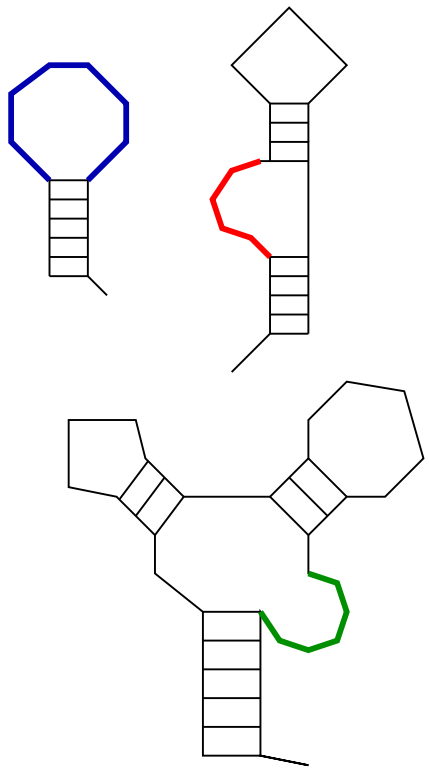
- Sfold by Ding & Lawrence
- Internal stability profile of small RNAs (5' and 3' end, average internal stability, . . .)
- probability profiling for prediction of unstructured regions in the target RNA:
  - generate a representative statistical sample of target RNA structures
  - calculate the probability that w consecutive positions, starting at position i, are within an unstructured region

2. Target accessibility and duplex thermodynamics

- Sfold by Ding & Lawrence
- Internal stability profile of small RNAs (5' and 3' end, average internal stability, . . .)
- probability profiling for prediction of unstructured regions in the target RNA:
  - generate a representative statistical sample of target RNA structures
  - calculate the probability that w consecutive positions, starting at position i, are within an unstructured region

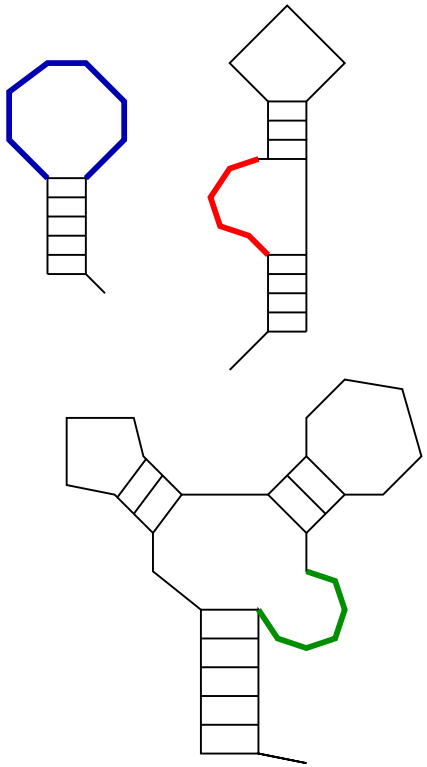
Complete ensemble of secondary structures

Probability of an unstructured region

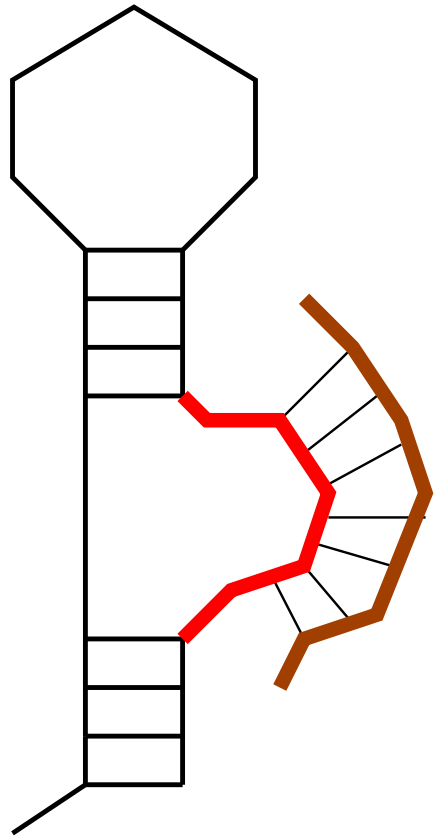


Complete ensemble of secondary structures

Probability of an unstructured region



Probability of interaction in unstructured regions



Outline

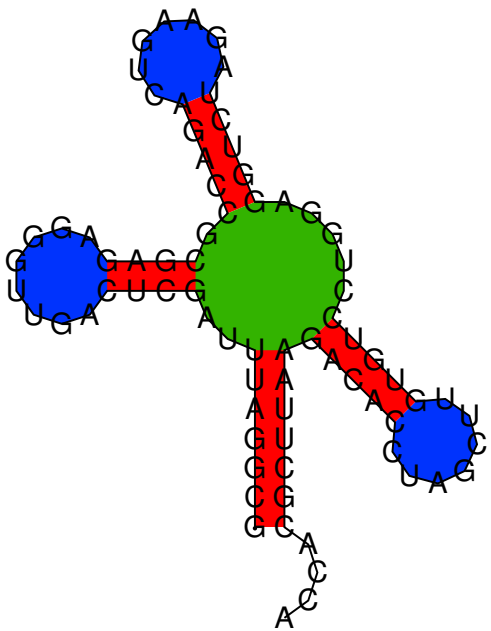
Basic Algorithms

- Mc Caskill: Equilibrium partition function for RNA secondary structure
- ViennaRNA Package

Basic Algorithms

- Mc Caskill: Equilibrium partition function for RNA secondary structure
- ViennaRNA Package

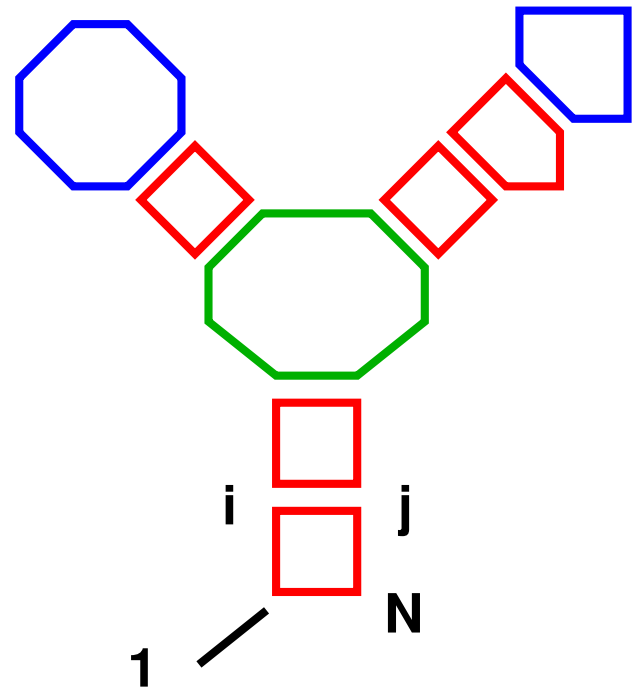
RNA Secondary Structure



- Hairpin Loops
- Interior Loops
- Multiloops
- Exterior Loops

Loop decomposition of RNA Secondary Structure

- $F(S) = \sum_{L \in S} F_L$.



Probability of a given secondary structure

- $Q = \sum_S e^{-[F(S)/kT]}$
- $P(S) = \frac{1}{Q} e^{-[F(S)/kT]}$

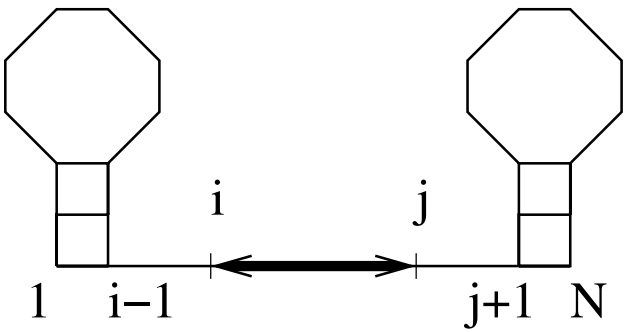
Probability of a given secondary structure

- $Q = \sum_S e^{-[F(S)/kT]}$
- $P(S) = \frac{1}{Q} e^{-[F(S)/kT]}$

Outline

$Pr_u[i,j]$: region i,j contains no secondary structure

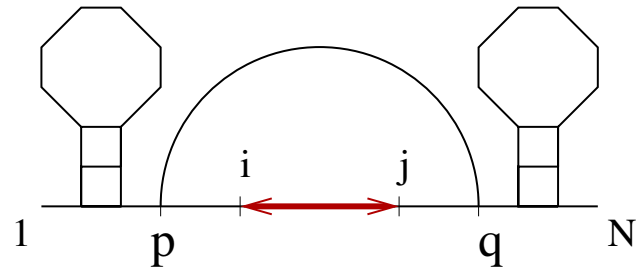
trivial case: region i,j is exterior to all loops.



$$Pr_u[i,j] = \frac{Q[1,i-1]Q[j+1,N]}{Q[1,N]}$$

$Pr_u[i,j]$: region i,j contains no secondary structure

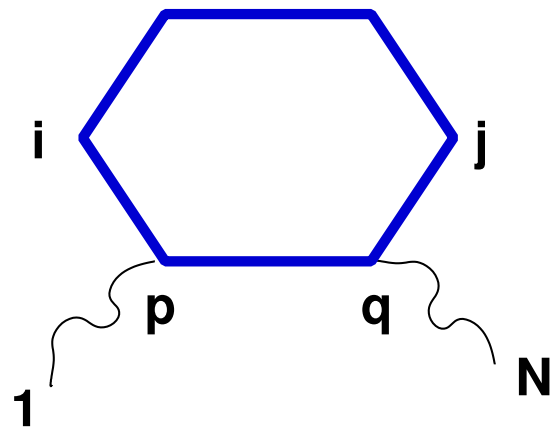
region i,j is spanned by a basepair (p,q).



$$Pr_u[i,j] \;\; += \sum_{\substack{p<i \\ j<q}} Prob[p,q] \frac{Qpq_u[i,j]}{Qb[p,q]}.$$

$Qpq_u[i,j]$: unpaired region i,j enclosed by pair (p,q):

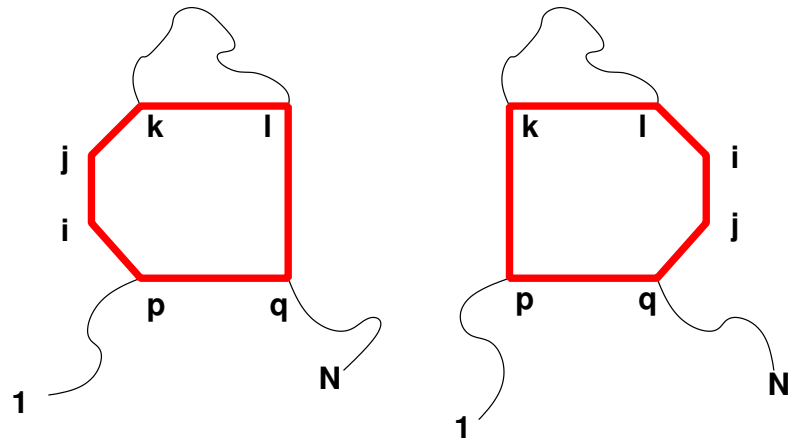
base pair (p,q) closes a Hairpin loop:



$$Qpq_u[i,j] = e^{-\beta H(p,q)} + \ldots$$

$Qpq_u[i,j]$: unpaired region i,j enclosed by pair (p,q):

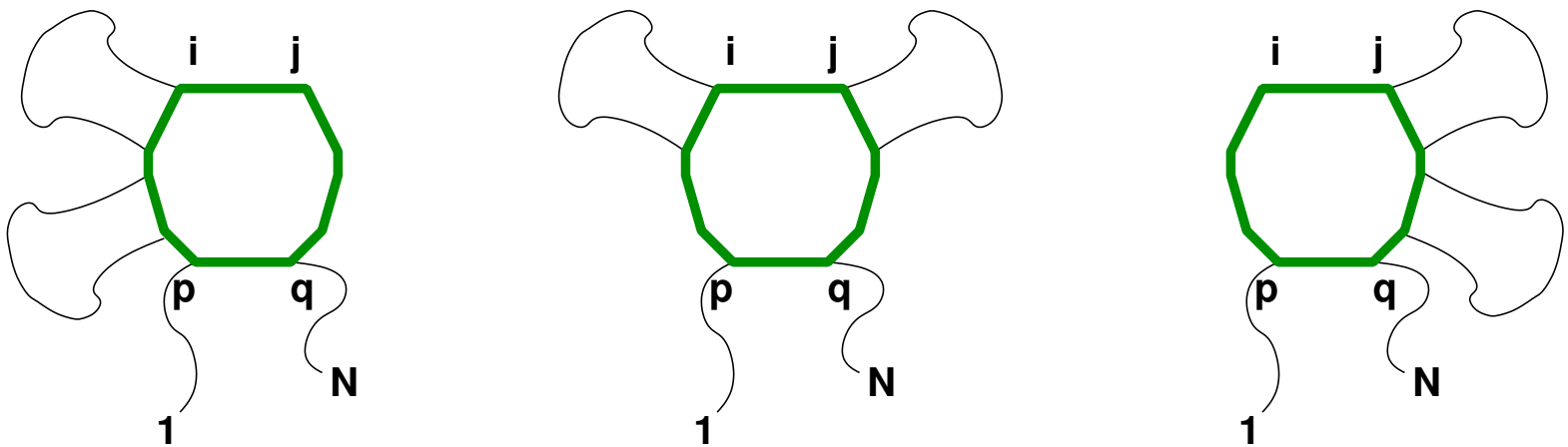
base pair (p,q) closes an Interior loop:



$$Qpq_u[i,j] += \sum_{\substack{p<i\leq j<k \\ l<i\leq j<q}} e^{-\beta I(p,q,k,l)} Qb[k,l] + \ldots$$

$Qpq_u[i,j]$: unpaired region i,j enclosed by pair (p,q):

base pair (p,q) closes a Multiloop:



$$Qpq_u[i,j] += \sum_{p<i\leq j<q}$$

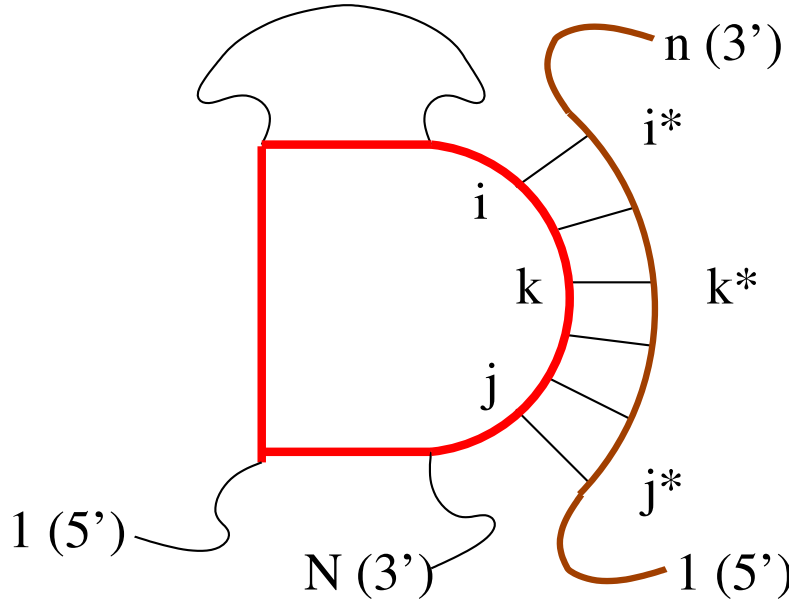
$$\Big(Qm2[p+1,i-1]e^{-\beta u(j-i+1)} + Qm[p+1,i-1]e^{-\beta u(j-i+1)}Qm[j+1,q-1] + e^{-\beta u(j-i+1)}Qm2[j+1,q-1]\Big).$$

$Pr_u[i,j]$: region i,j contains no secondary structure

$$Pr_u[i,j] = \frac{Q[1,i-1]Q[j+1,N]}{Q[1,N]} + \sum_{p<i<j<q} Prob[p,q]\frac{Qpq_u[i,j]}{Qb[p,q]}.$$

Outline

unstructured region i, j is paired to i^*, j^*:



$$Qup[i,j,i^*,j^*] = Pr_u[i,j] \sum_{\substack{i<k<j \\ i^*>k^*>j^*}} Qp[i,k,i^*,k^*] e^{-\beta I(k,k^*;j,j^*)}.$$

unstructured region i, j is paired to i^*, j^*:

$$Pup[i, j, i^*, j^*] = \frac{Qup[i, j, i^*, j^*]}{Qup}.$$

where

$$Qup = \sum_{i,\ j,\ i^*, j^*} Qup[i, j, i^*, j^*].$$

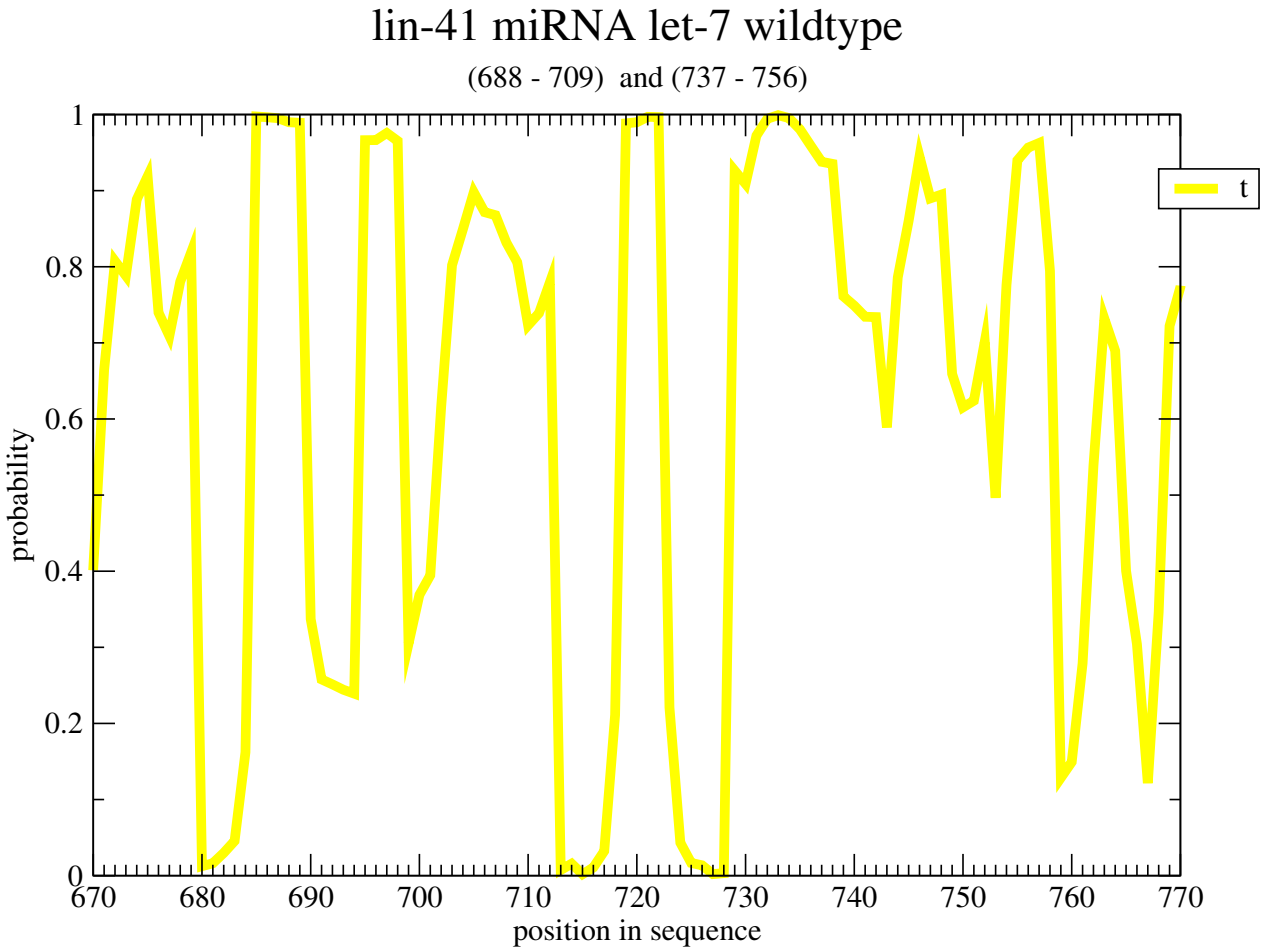
position i is contained in a target site for a small RNA

$$Pup[i,j] = \sum_{i^*,j^*} Pup[i,j,i^*,j^*].$$

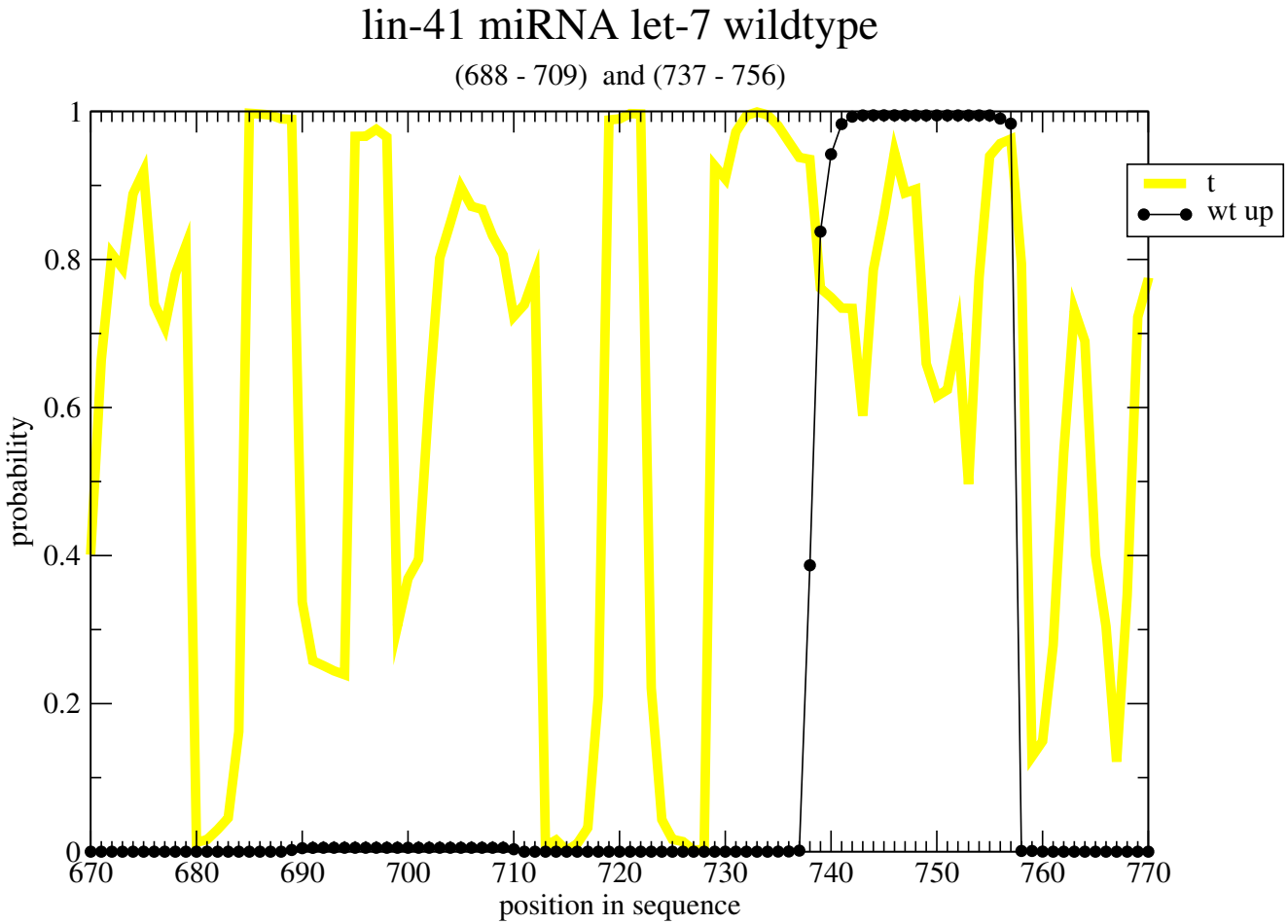
and

$$Pup[k] = \sum_{i \leq k \leq j} Pup[i,j].$$

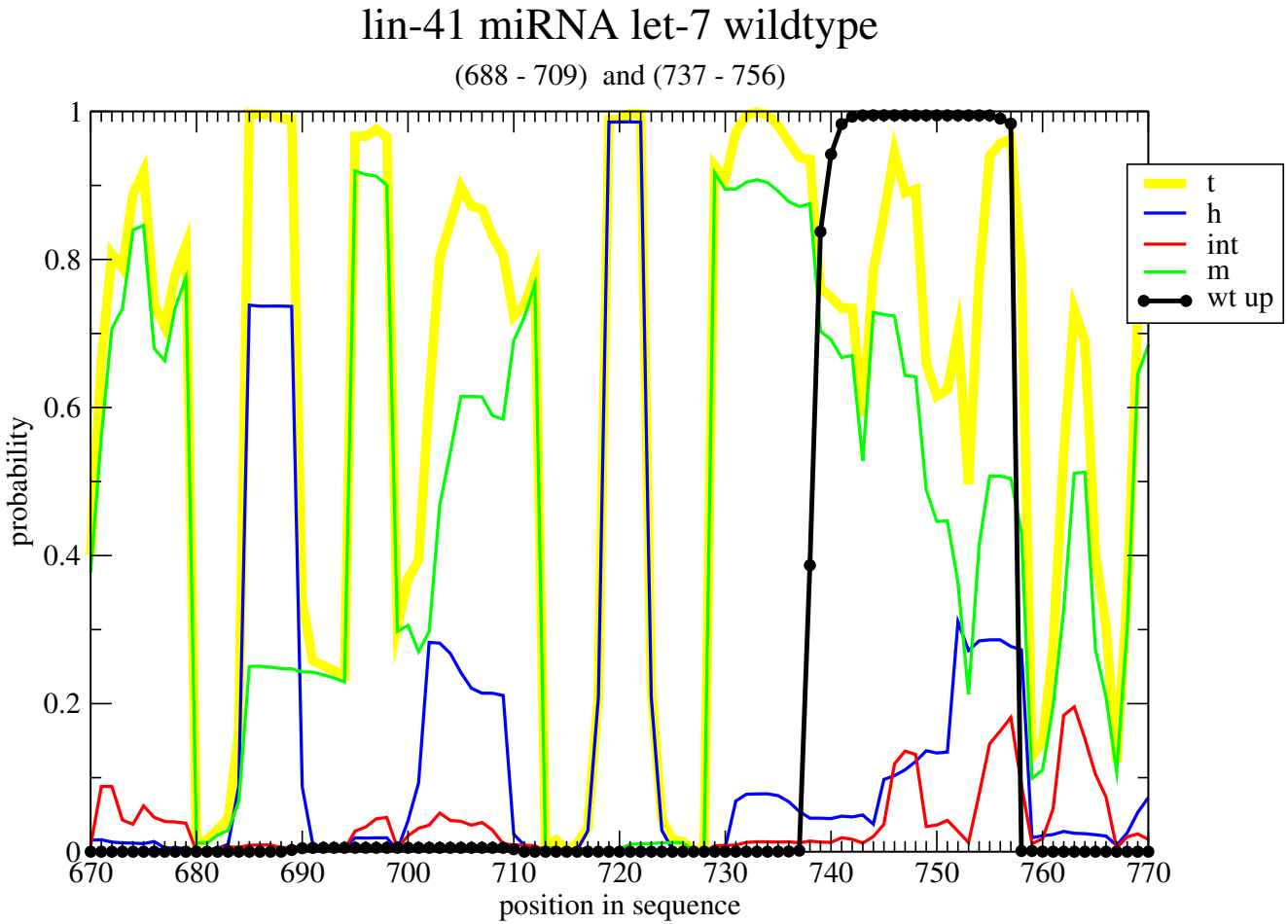
Interaction of let-7 miRNA with target gene lin-1



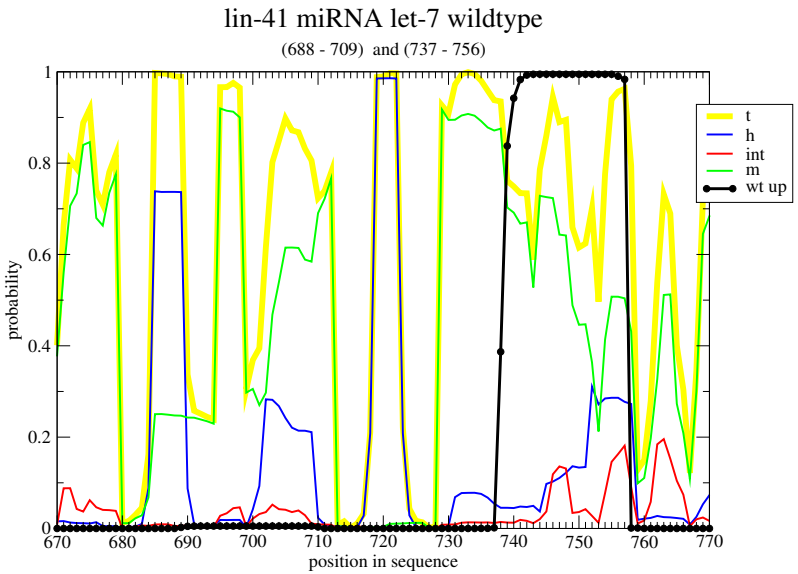
Interaction of let-7 miRNA with target gene lin-1

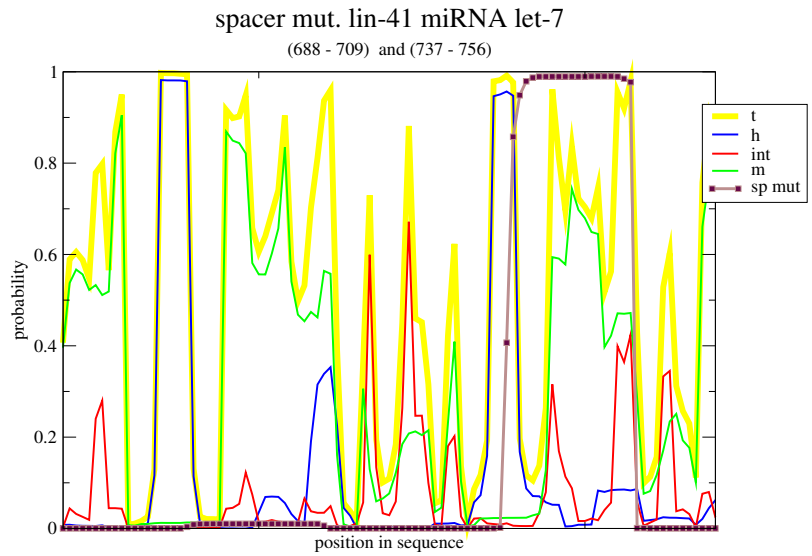


Interaction of let-7 miRNA with target gene lin-1

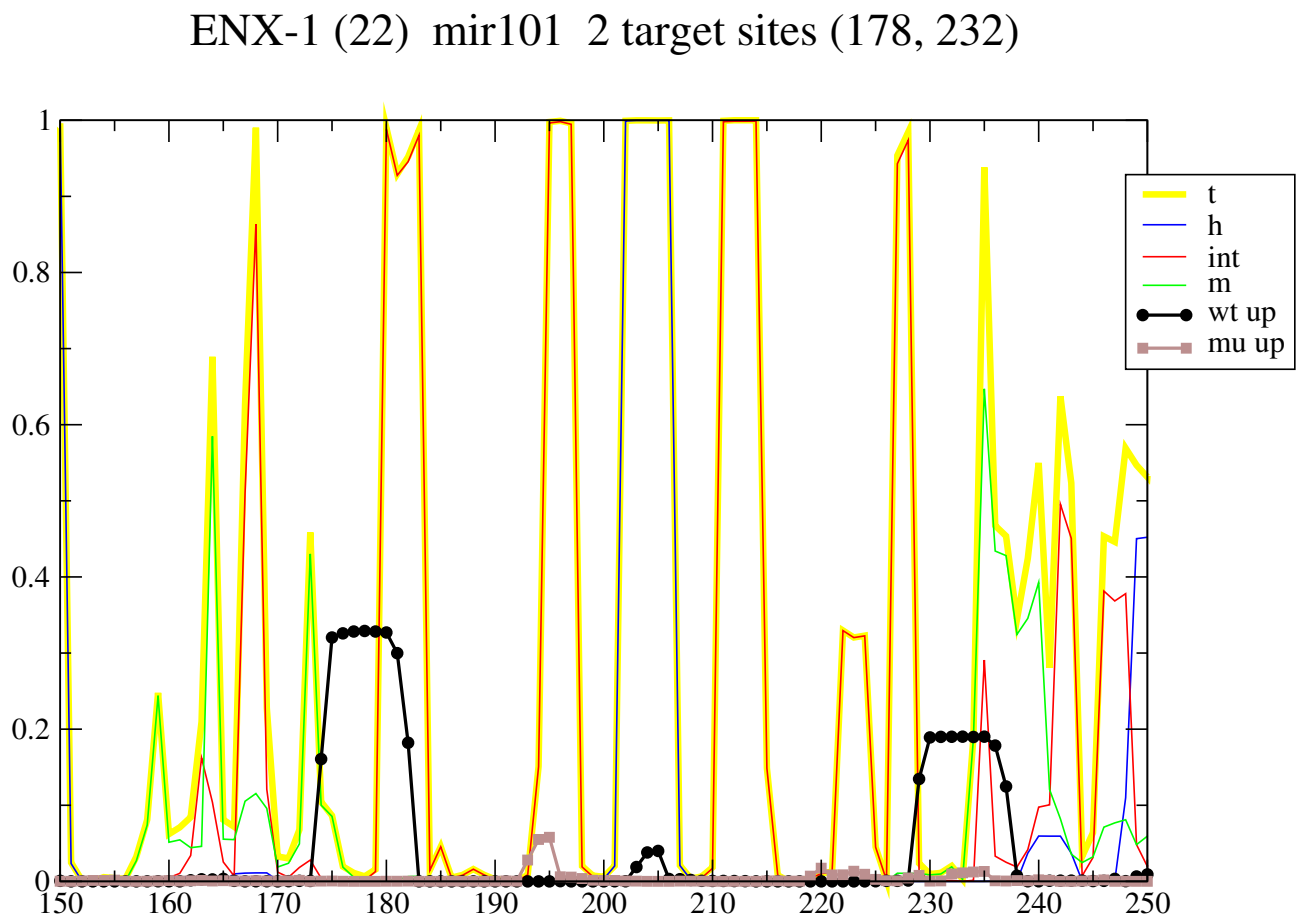


Interaction of let-7 miRNA with target gene lin-1



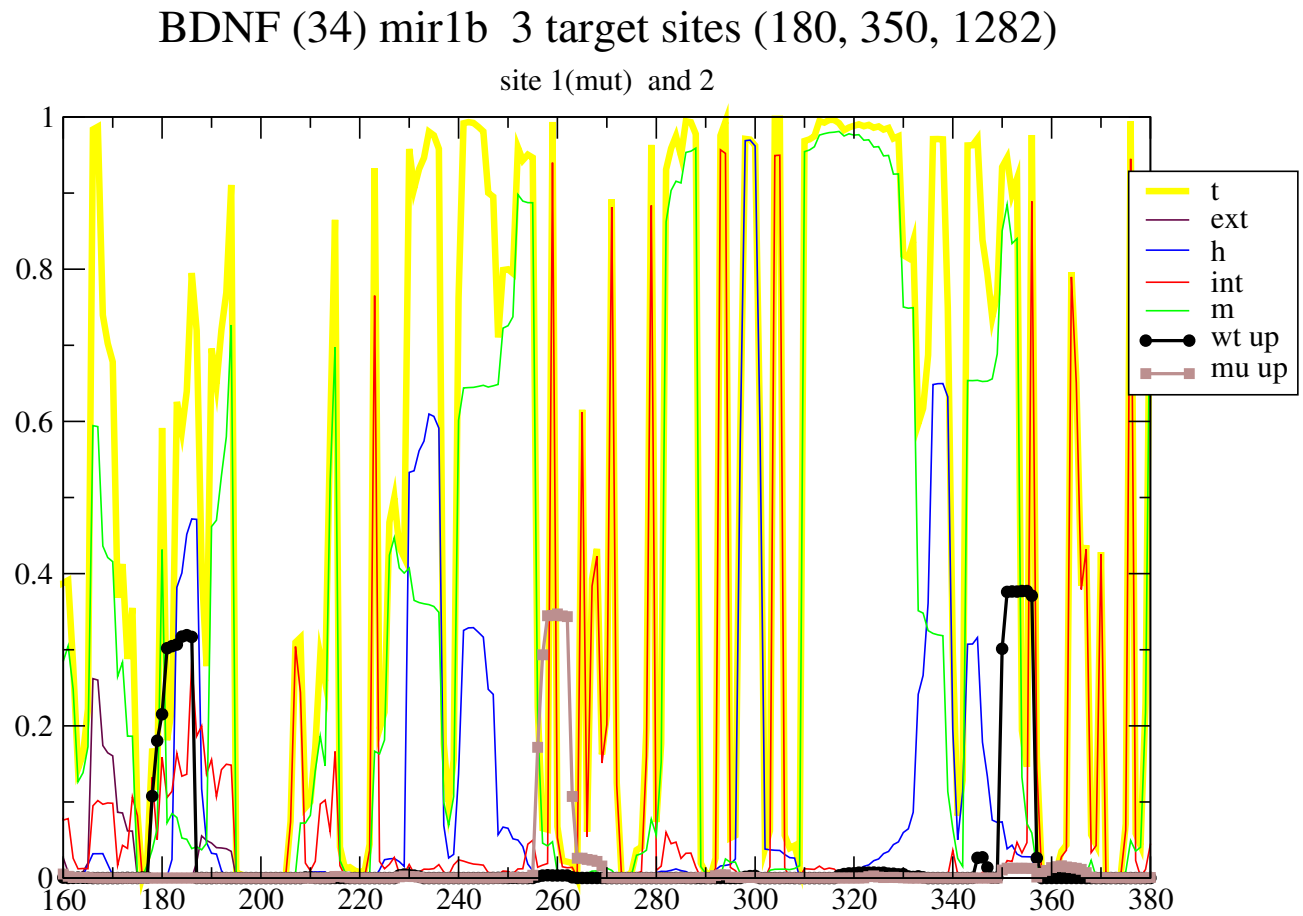


Interaction of human enx-1 with miRNA mir101



Interaction of human bdnf with miRNA mir1b

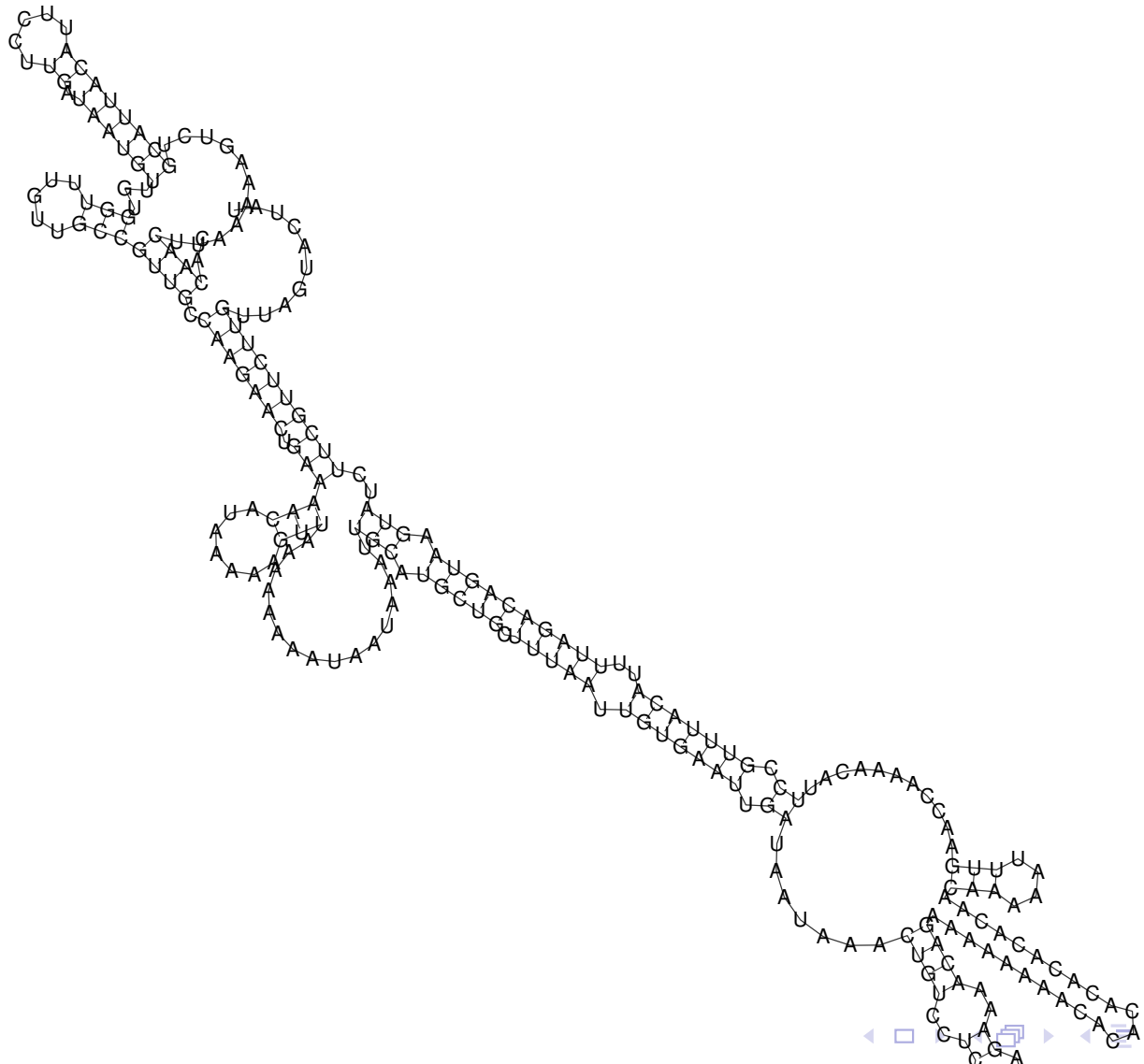
176-185 p=0.99, 348-358, p<0.01



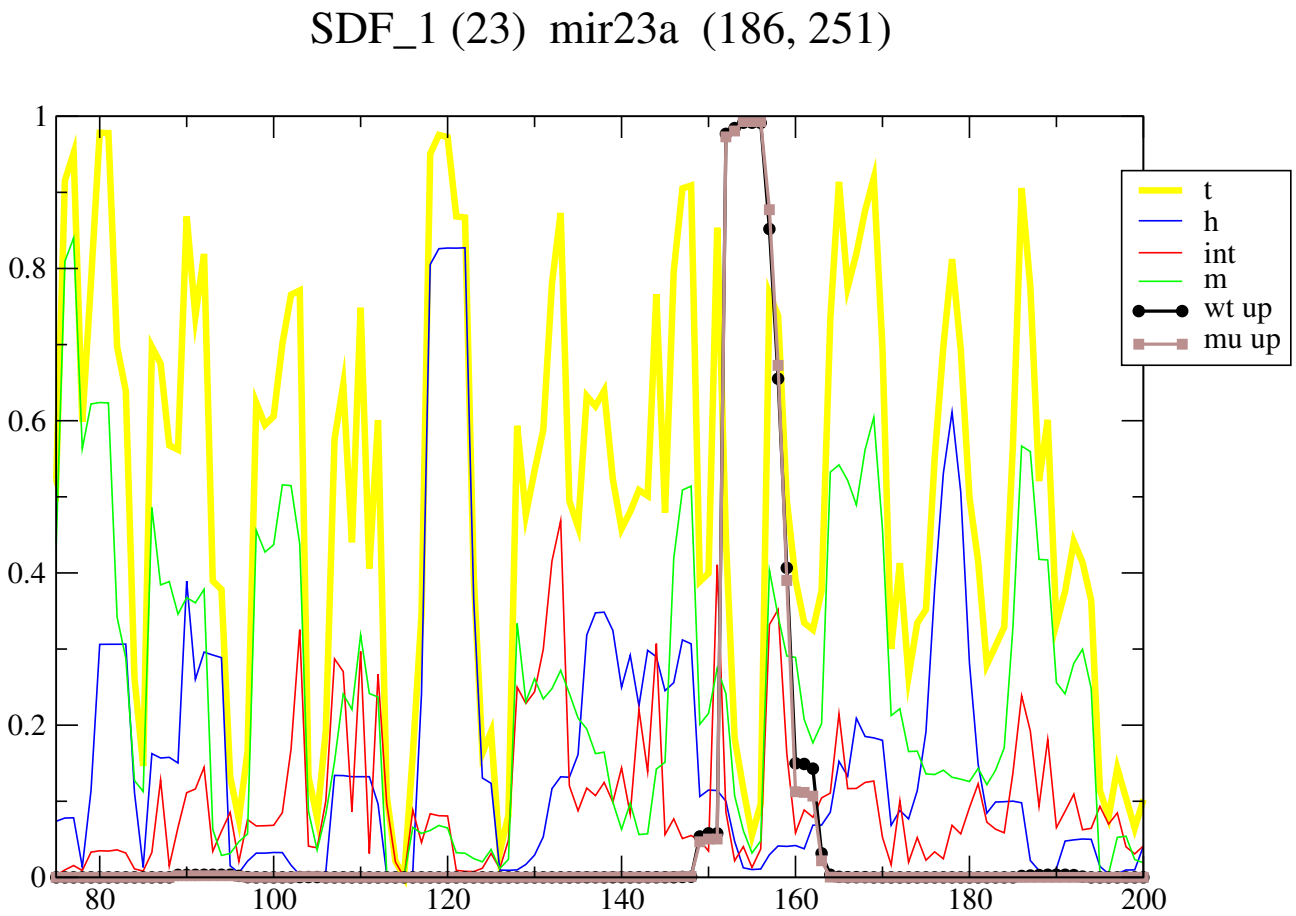
Interaction of human bdnf with miRNA mir1b

176-185 p=0.99, 348-358, p<0.01

seed sequence: ACATTCC

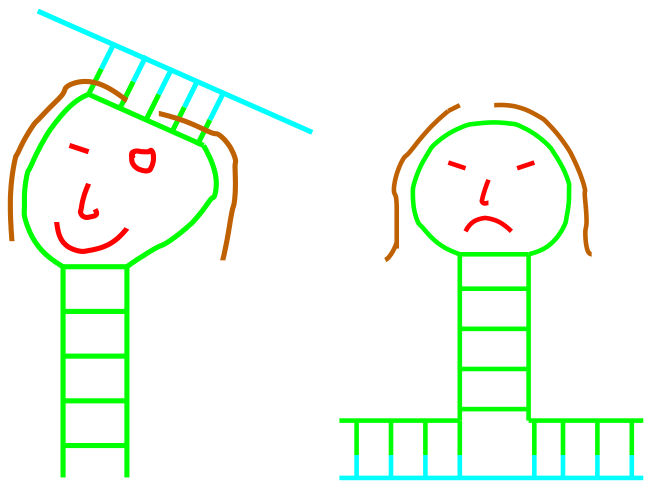


Interaction of human sdf-1 with miRNA mir23a

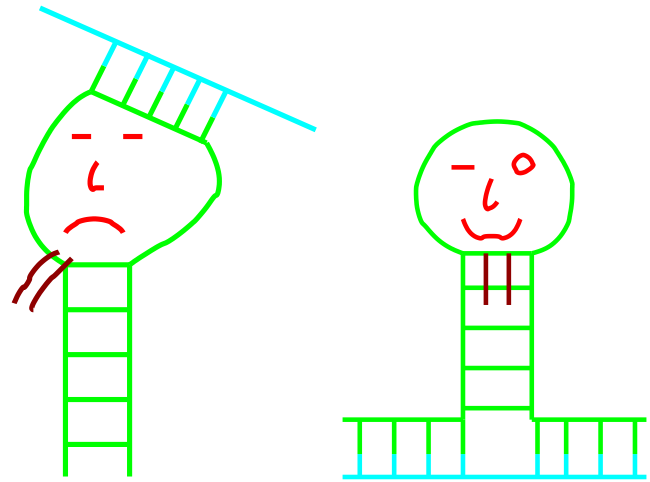


2 Parts of one Solution

RNAup



RNAcofold



Outline

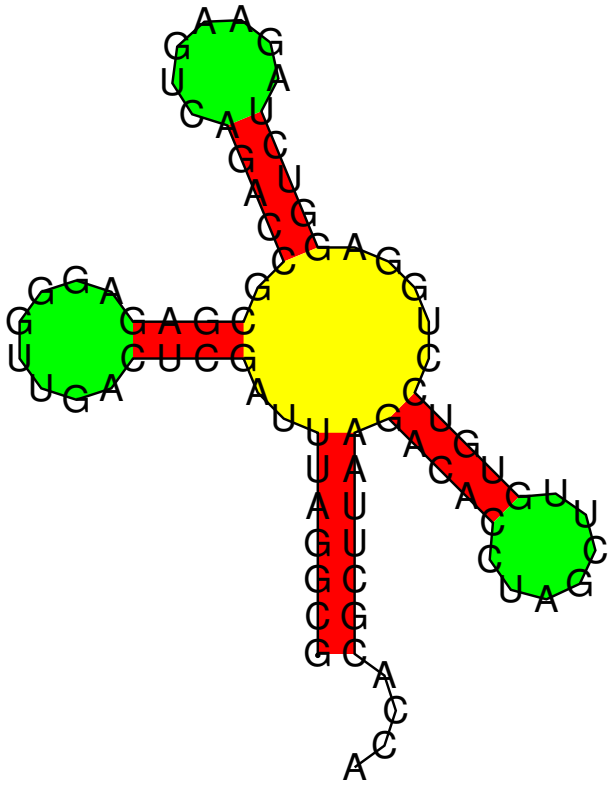
RNAcofold

- Variant of RNAfold
- Concatenate the 2 molecules
- Define cut between molecules

RNA Secondary Structure

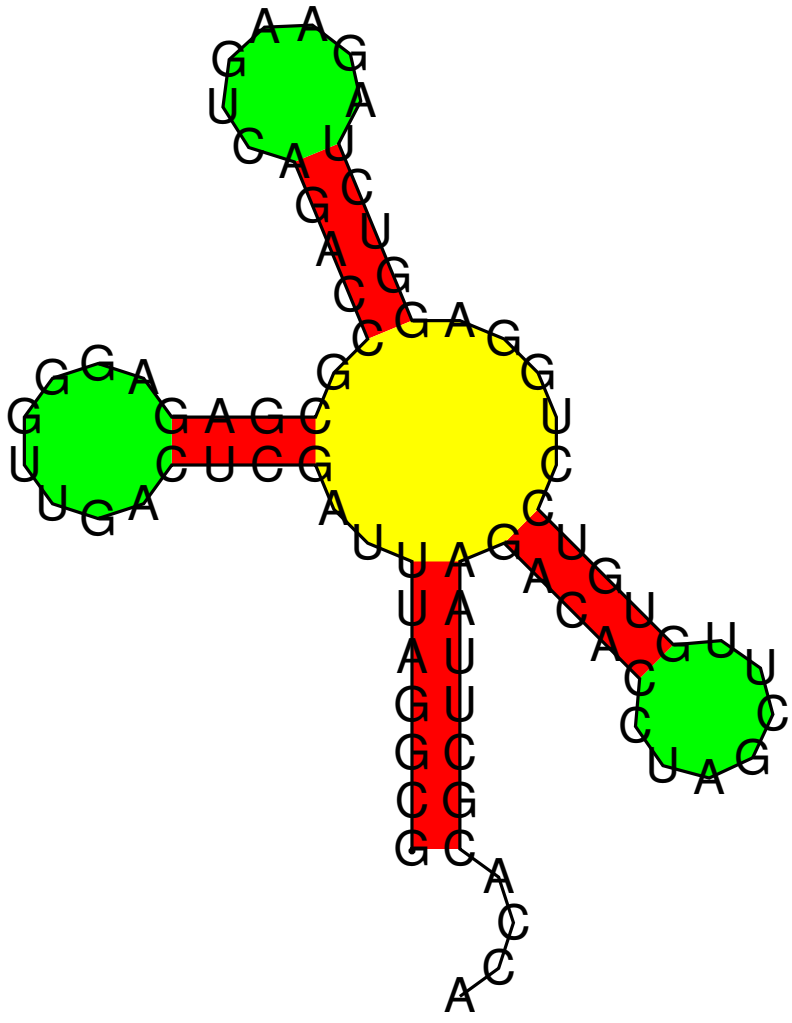
Free Energy

Free Energy is sum of Loop contributions:

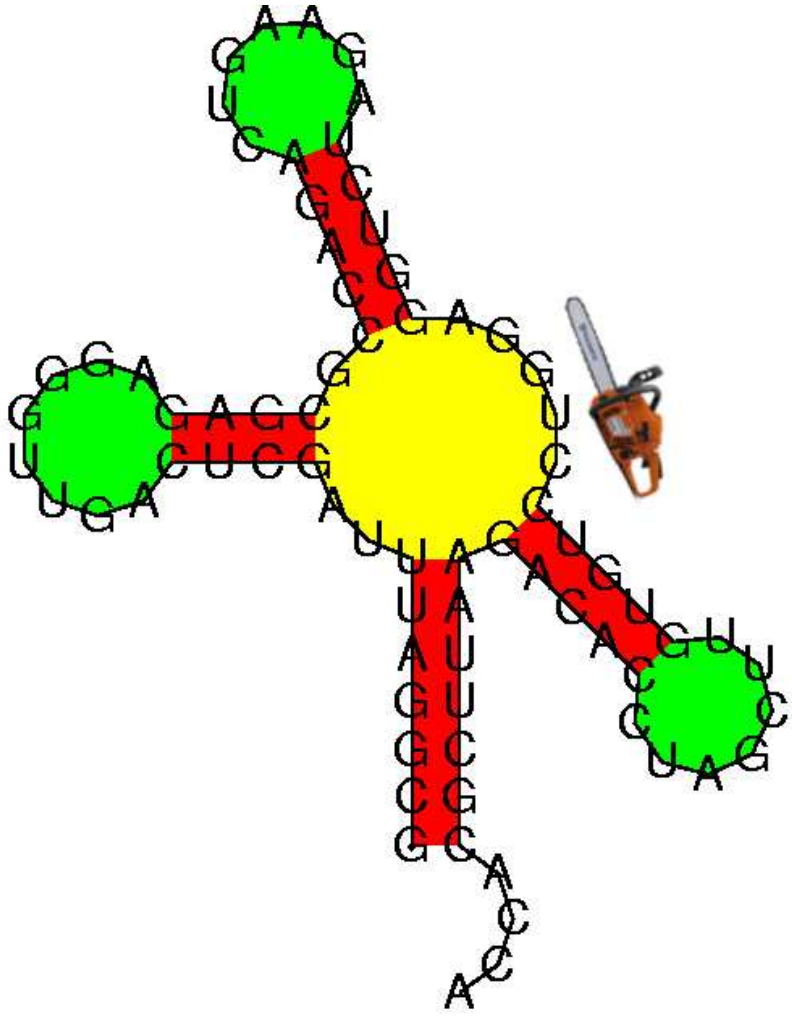


- Interior Loops
- Hairpin Loops
- Multi loops
- **Exterior Loops**

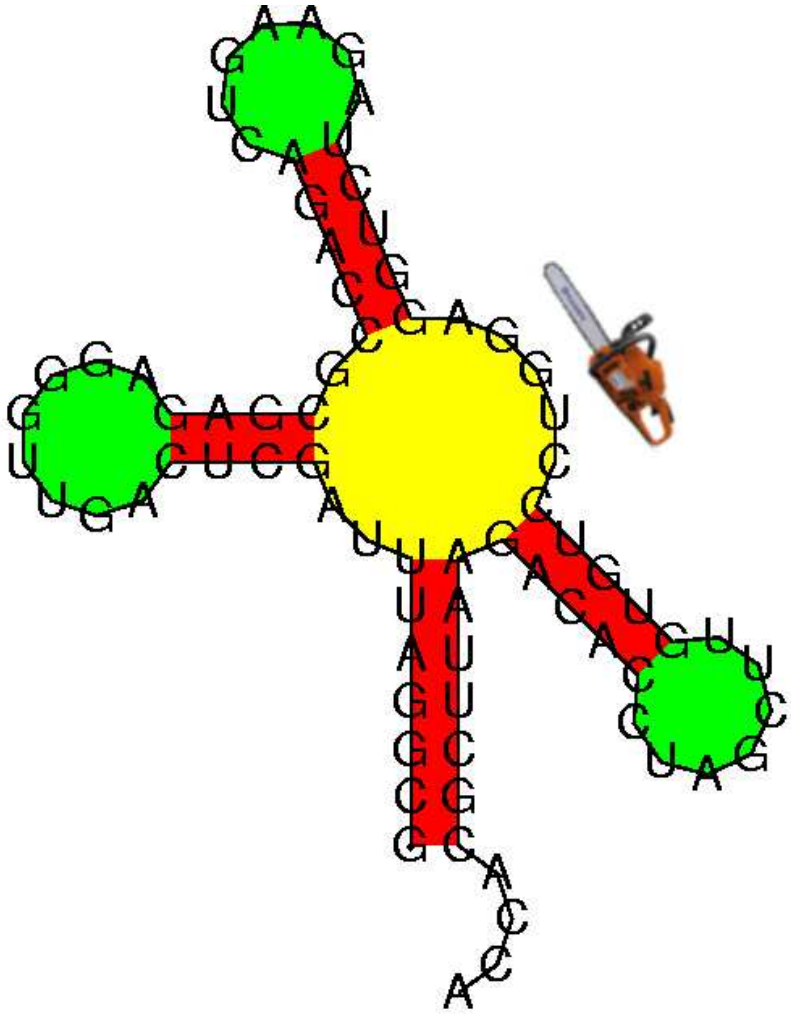
Going exterior



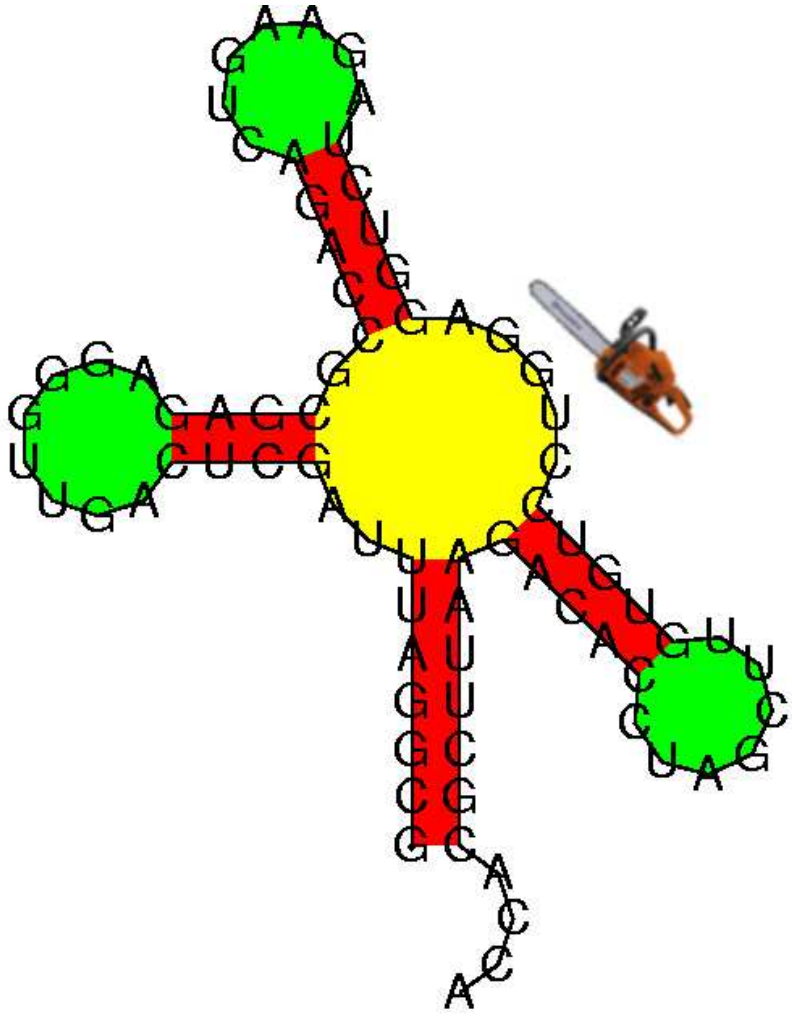
Going exterior



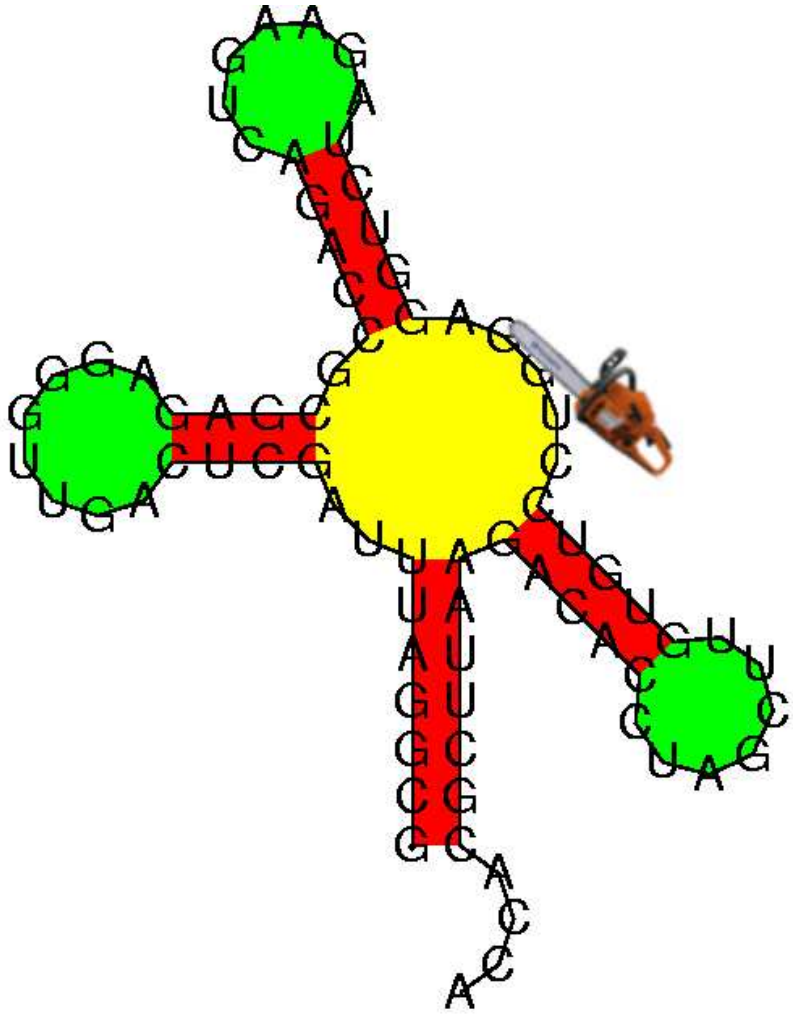
Going exterior



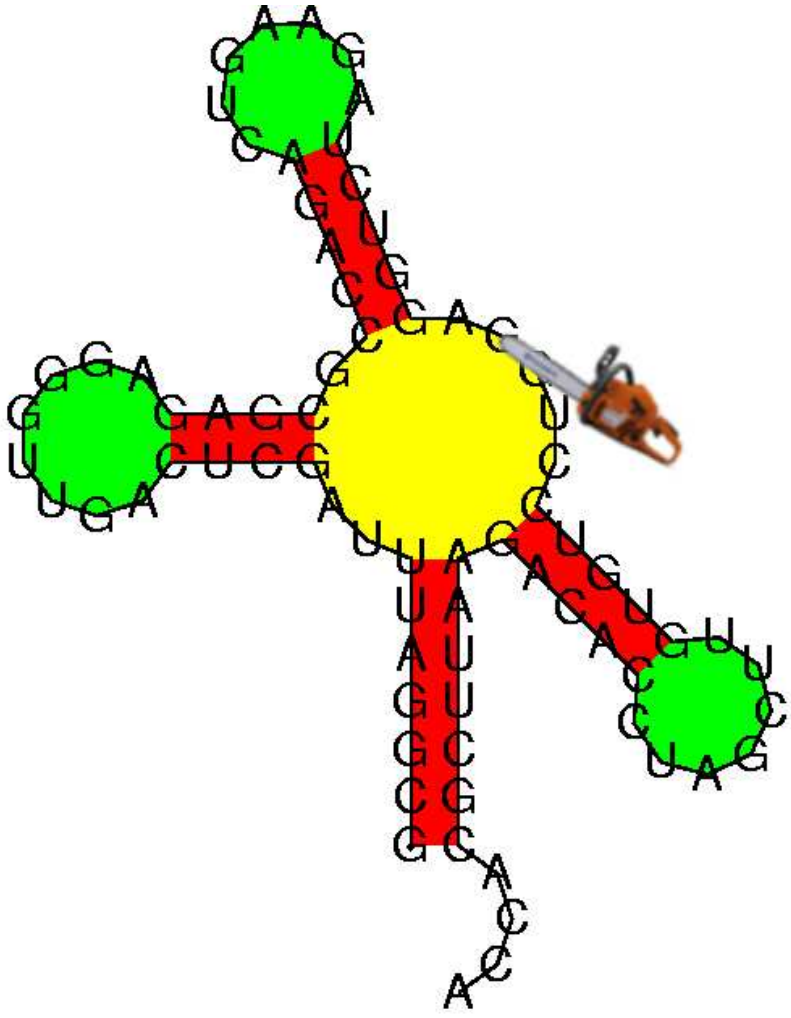
Going exterior



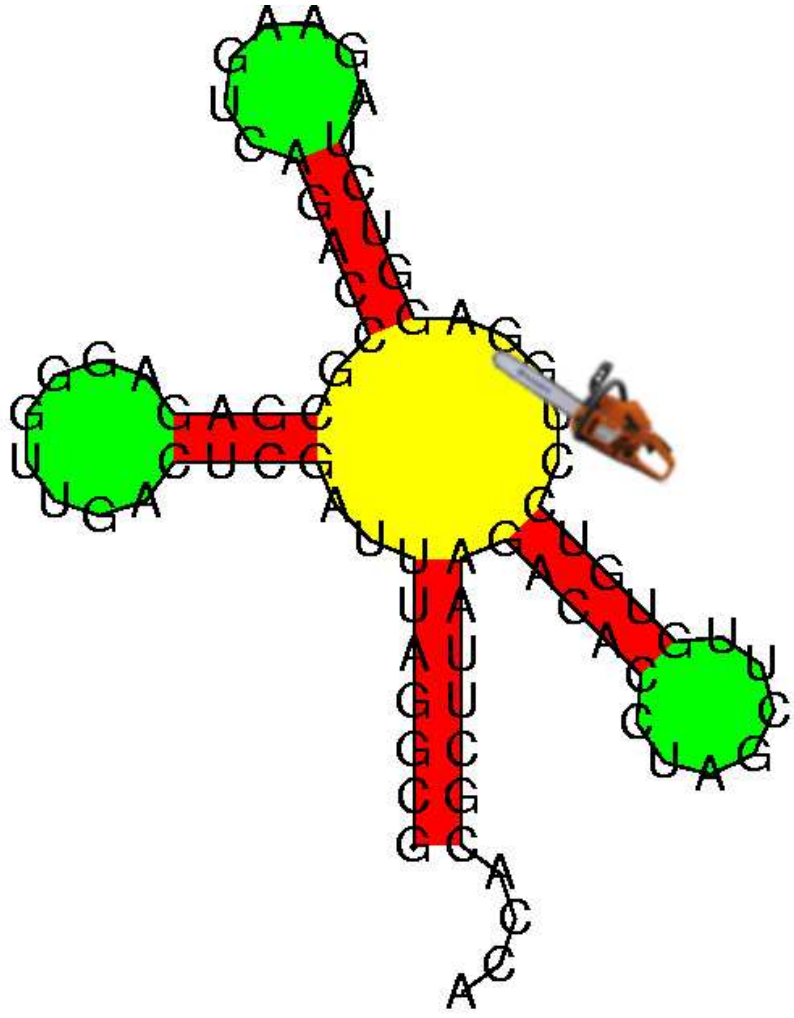
Going exterior



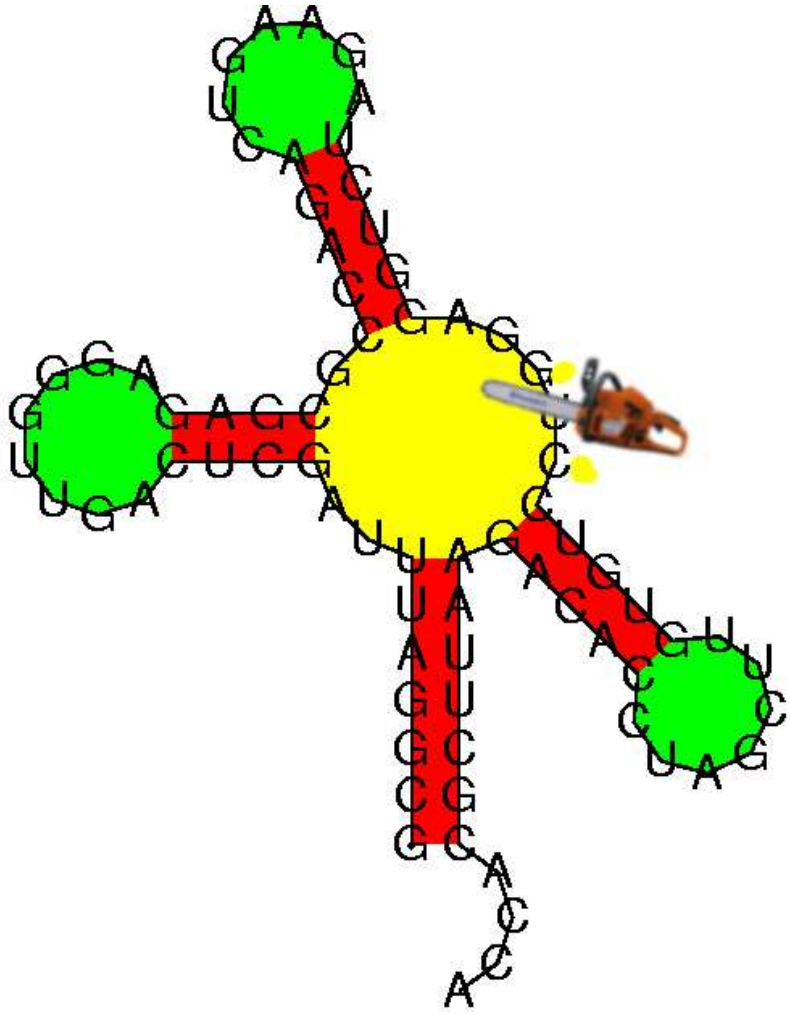
Going exterior



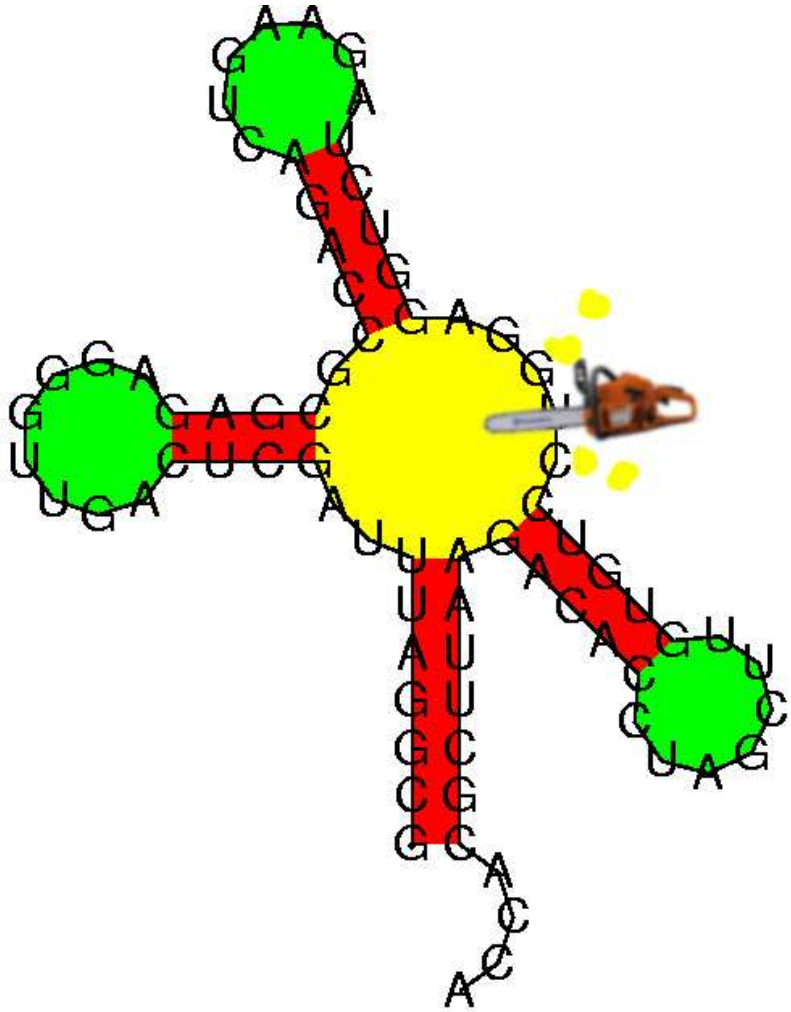
Going exterior



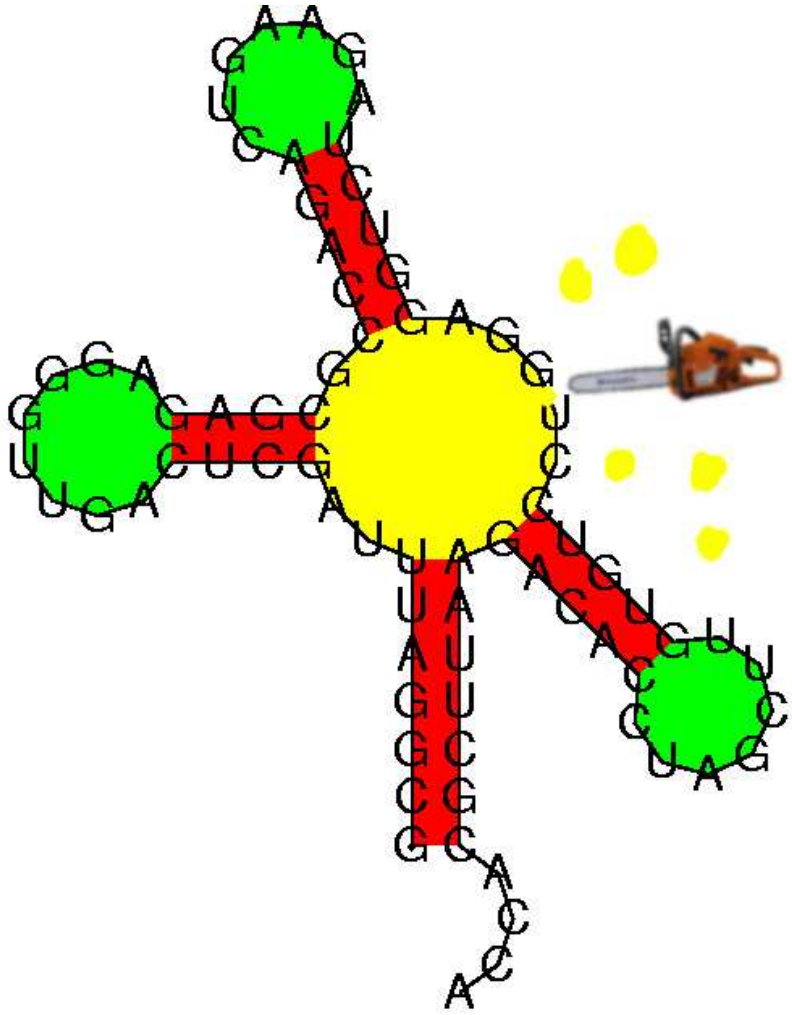
Going exterior



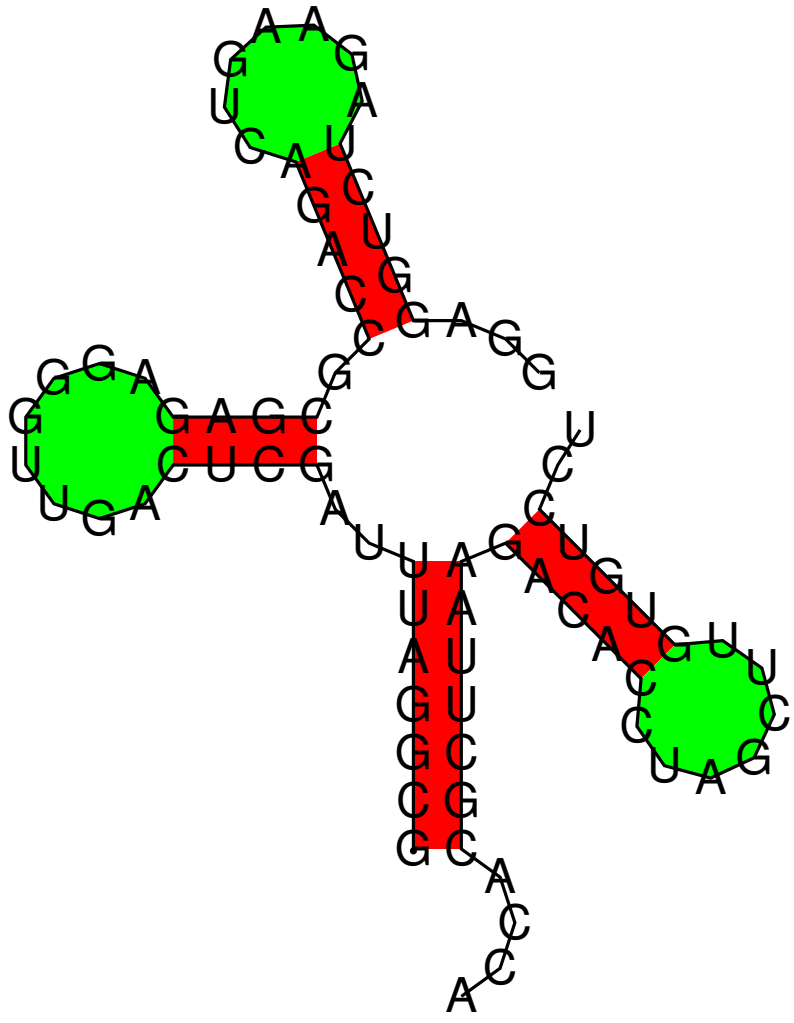
Going exterior



Going exterior



Going exterior



Partition Function

recursion

Partition function of segment i, j

$$Q_{ij} = Q_{i+1,j} + \sum_{i<k\leq j} Q^B_{i,k} Q_{k+1,j}$$

$$Q^B_{i,j} = \mathcal{H}(i,j) + \sum_{i<k<l<j} Q^B_{kl} \mathcal{I}(i,j;k,l) + Q^M_{i+1,j-1} a$$

$$Q^M_{i,j} = Q^M_{i+1,j} c + \sum_{i<k\leq j} Q^B_{i,k} b Q^{M1}_{k+1,j}$$

$$Q^{M1}_{i,j} = Q^{M1}_{i+1,j} + \sum_{i<k<j} Q^B_{i,k} b Q^{M1}_{k+1,j} + \sum_{i<k\leq j} Q^B_{i,k} b(j-k) c$$

Partition Function

recursion

Parts to be checked when cofolding

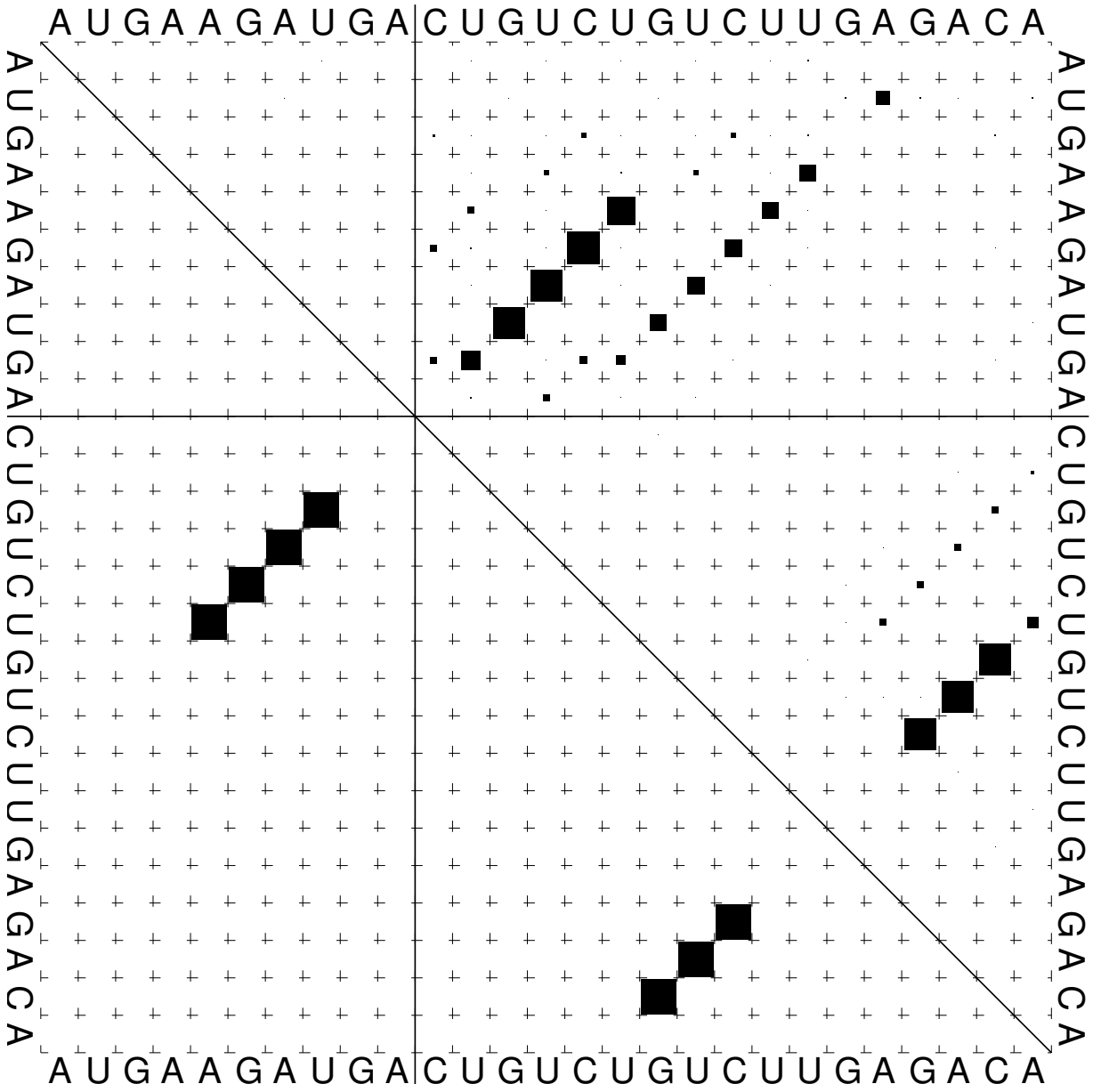
$$Q_{ij} = Q_{i+1,j} + \sum_{i<k\leq j} Q^B_{i,k} Q_{k+1,j}$$

$$Q^B_{i,j} = \mathcal{H}(i,j) \vee Q_{i+1,\|} Q_{\|,j-1} + \sum_{i<k<l<j} Q^B_{kl} \mathcal{I}(i,j;k,l) + Q^M_{i+1,j-1} a$$

$$Q^M_{i,j} = Q^M_{i+1,j} c + \sum_{i<k\leq j} Q^B_{i,k} b Q^{M1}_{k+1,j}$$

$$Q^{M1}_{i,j} = Q^{M1}_{i+1,j} + \sum_{i<k<j} Q^B_{i,k} b Q^{M1}_{k+1,j} + \sum_{i<k\leq j} Q^B_{i,k} b(j-k) c$$

Output: Dot Plot

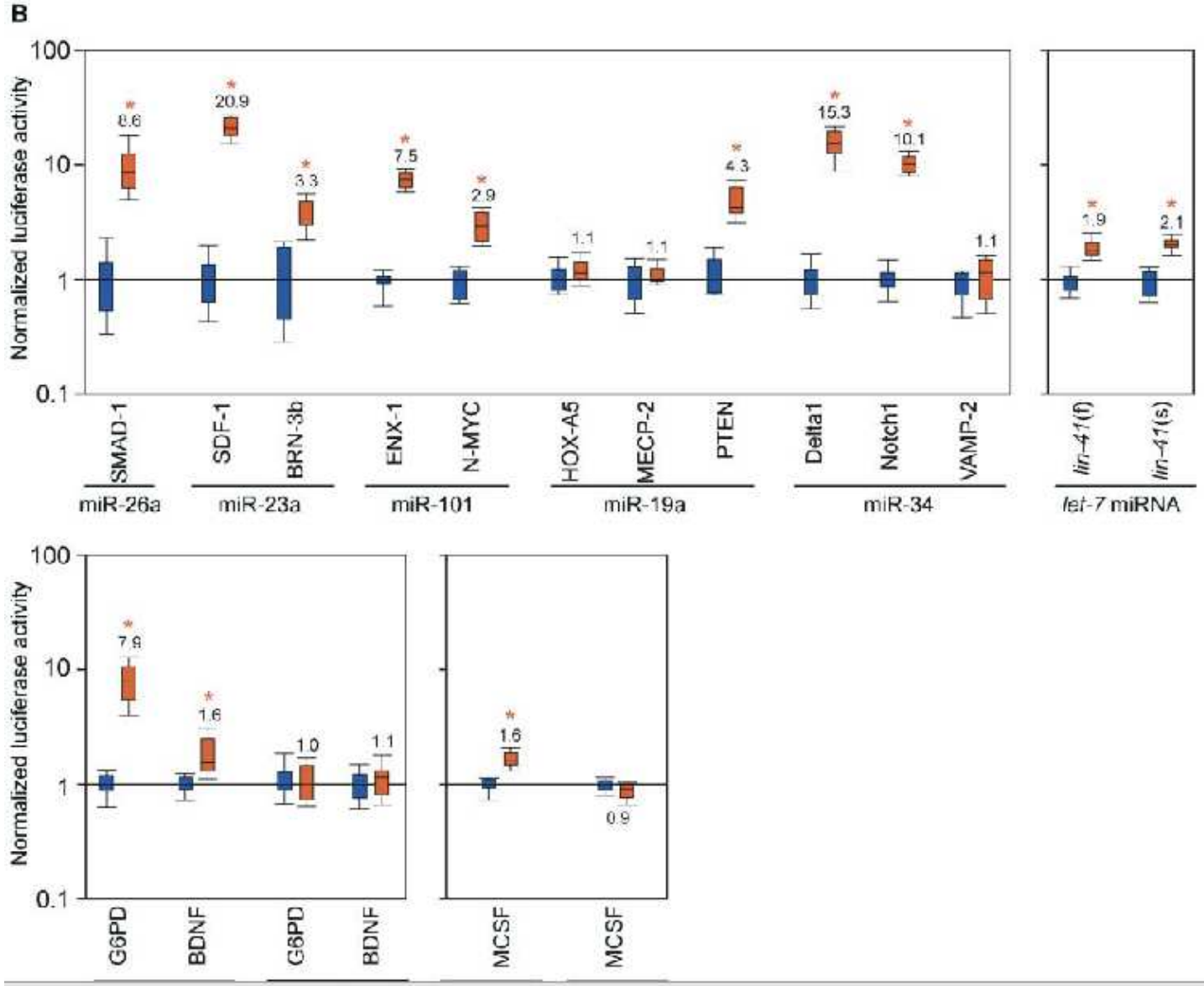


Outline

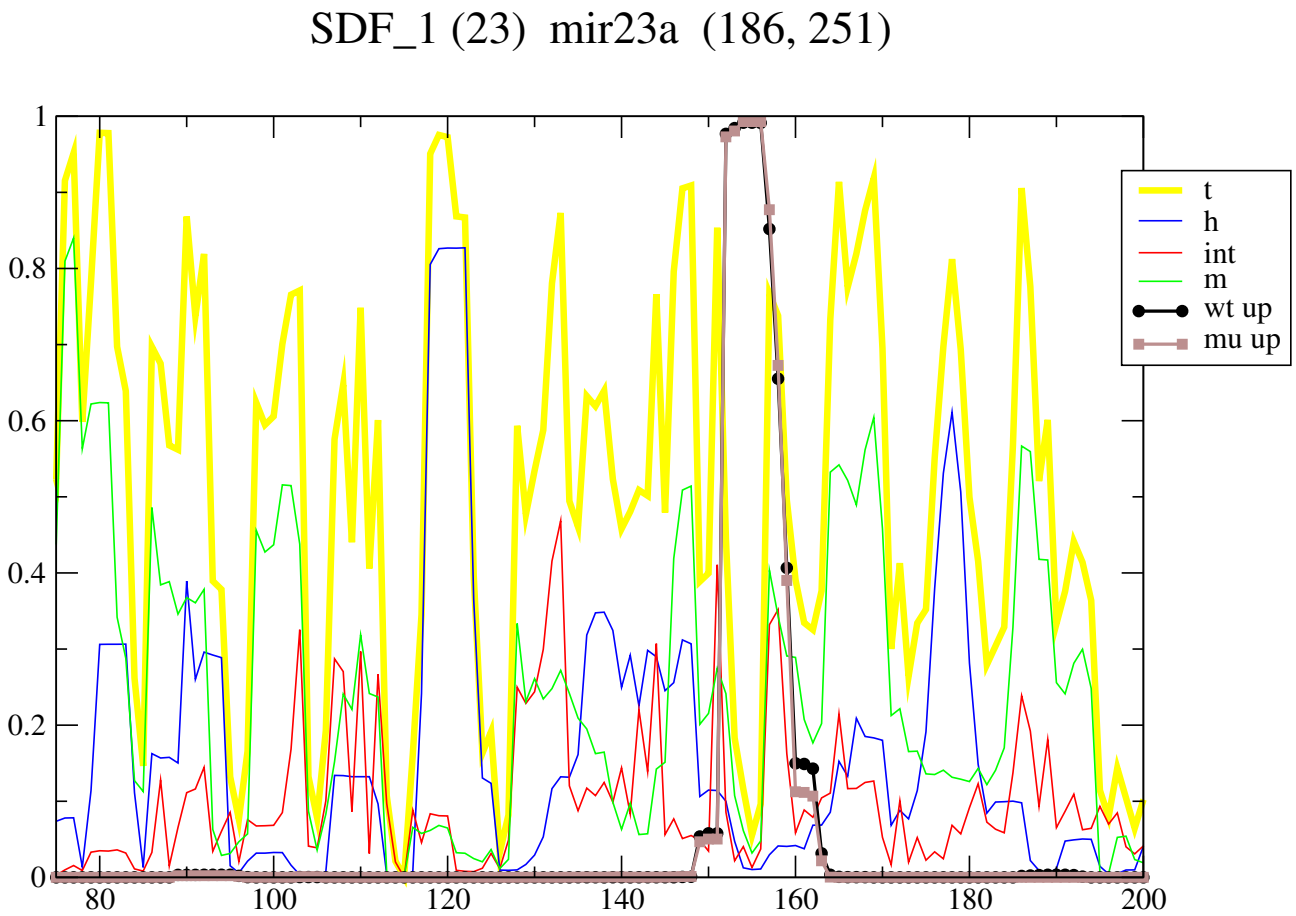
Setting

Wenn du nicht überzeugen kannst, stifte Verwirrung

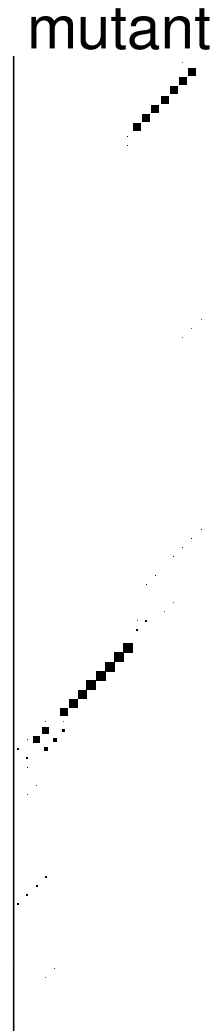
Lewis et.al, Prediction of Mammalian MicroRNA Targets



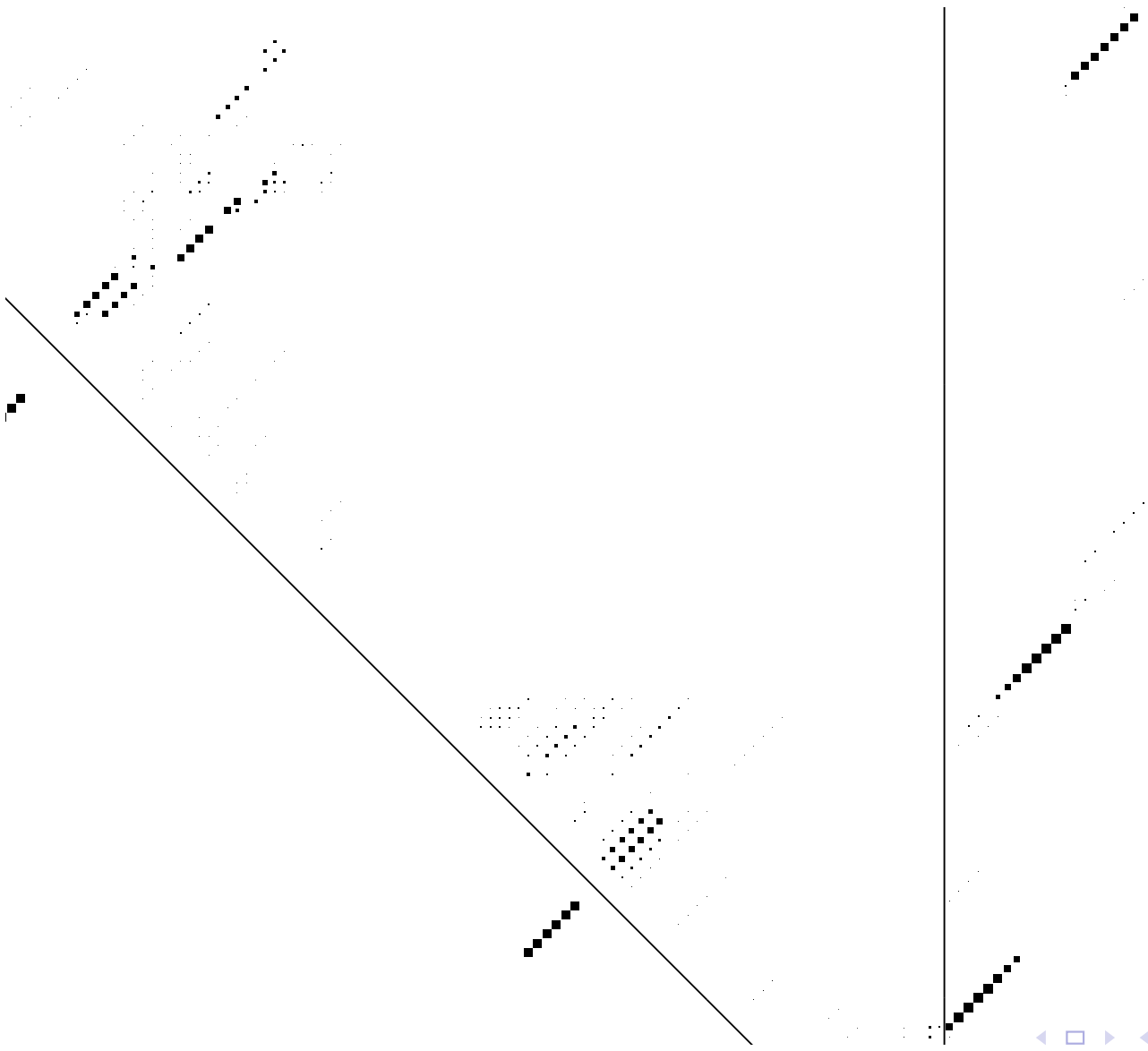
Interaction of human sdf-1 with miRNA mir23a



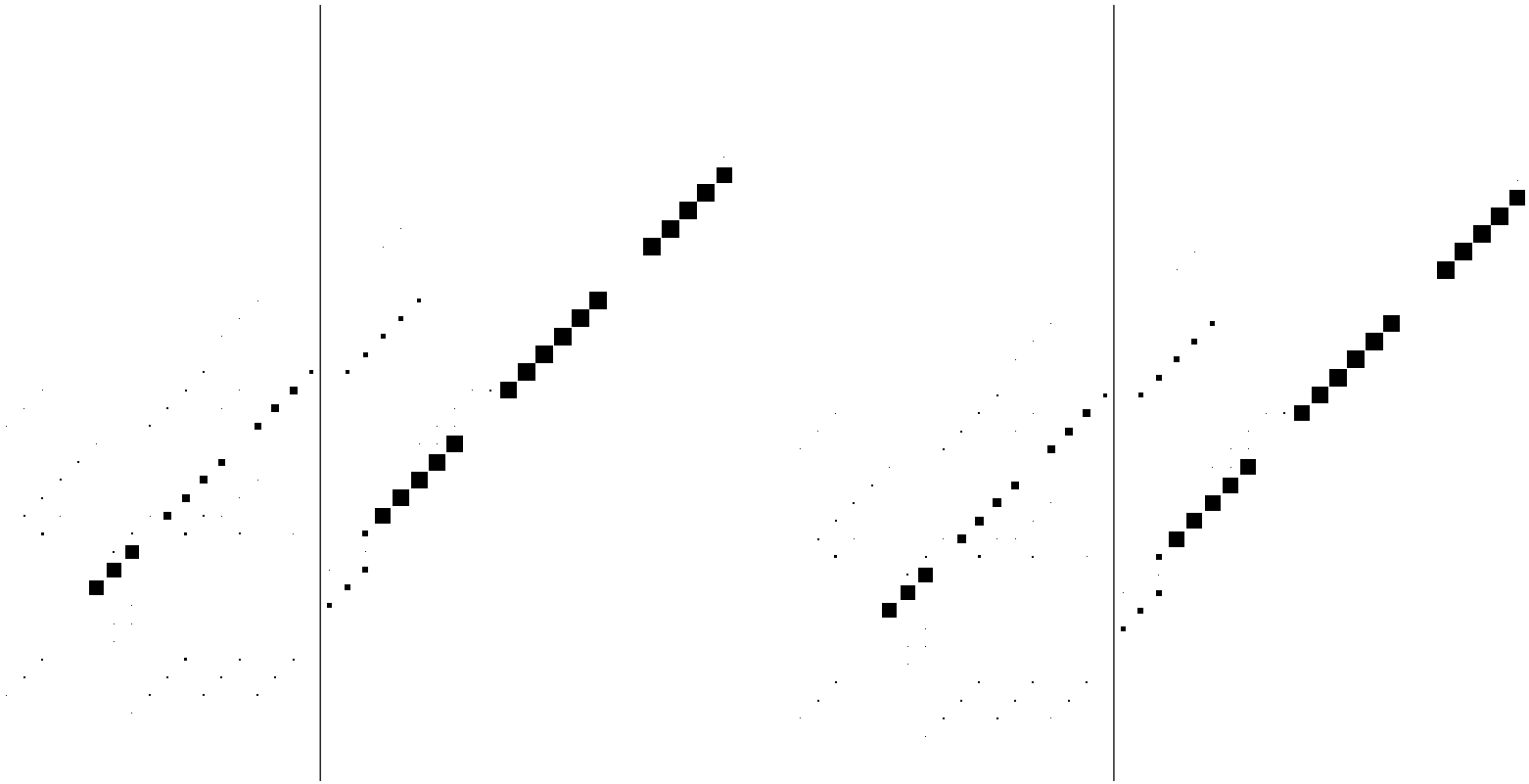
SDF1



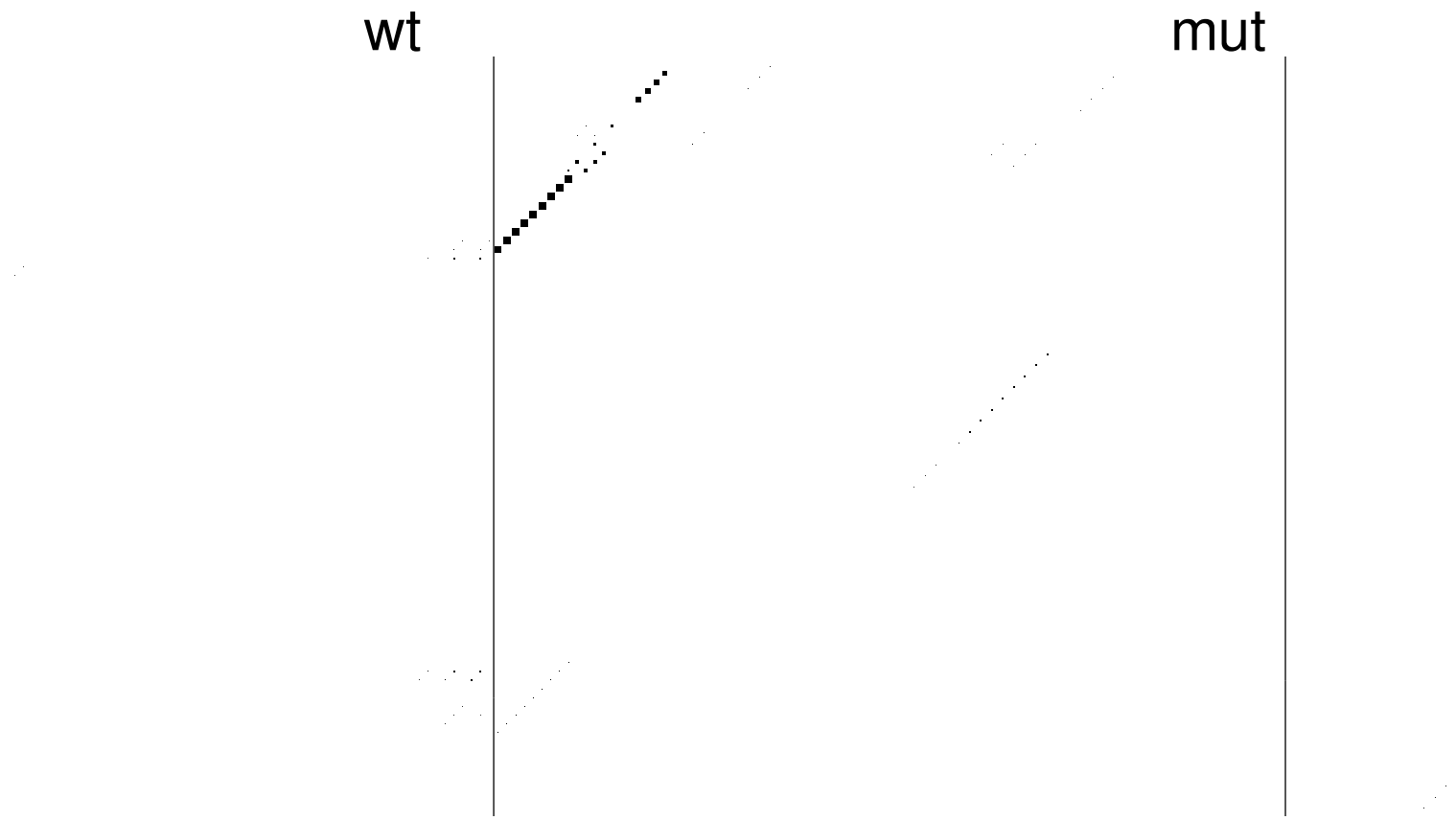
SDF1



HoxA5, position 185-230

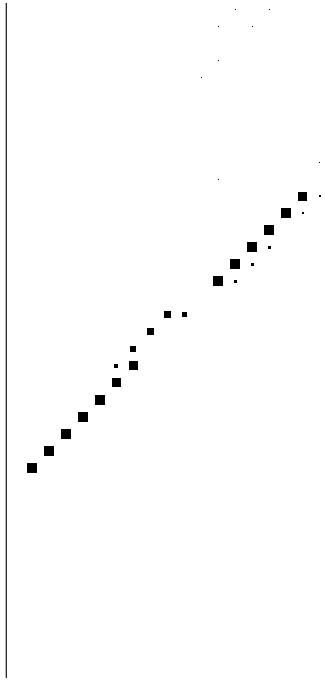


Enx1, position 150-235

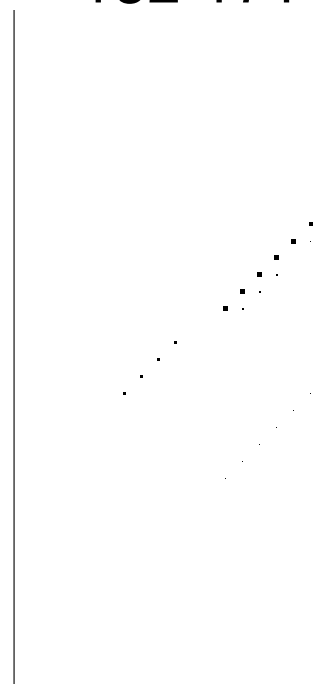


BRN3-b

wt
132-171, 30%



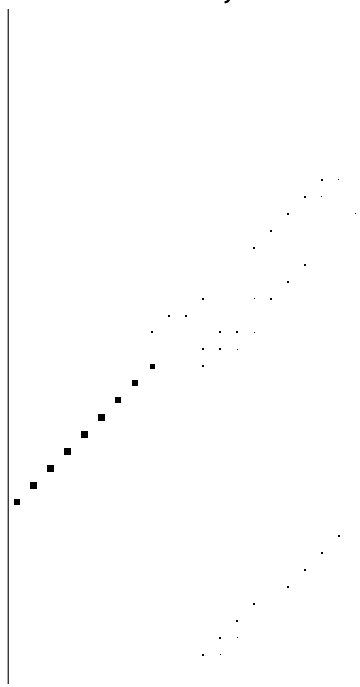
mutant
132-171



BRN3-b

wt

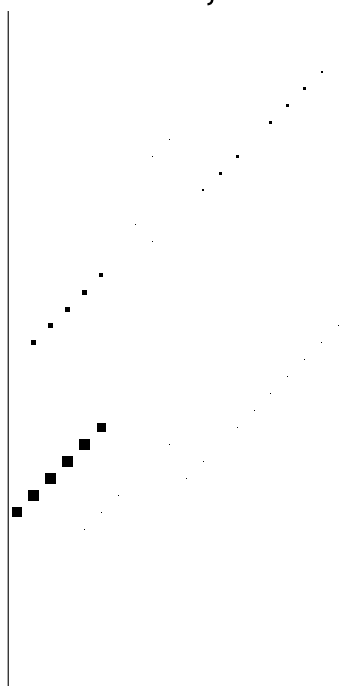
492-531, 12%



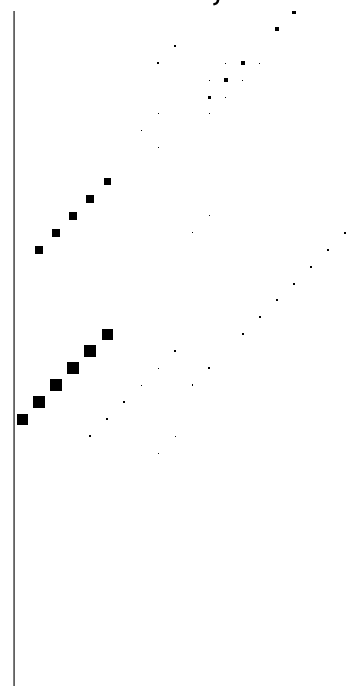
mutant

BRN3-b

wt
622-66, 40%



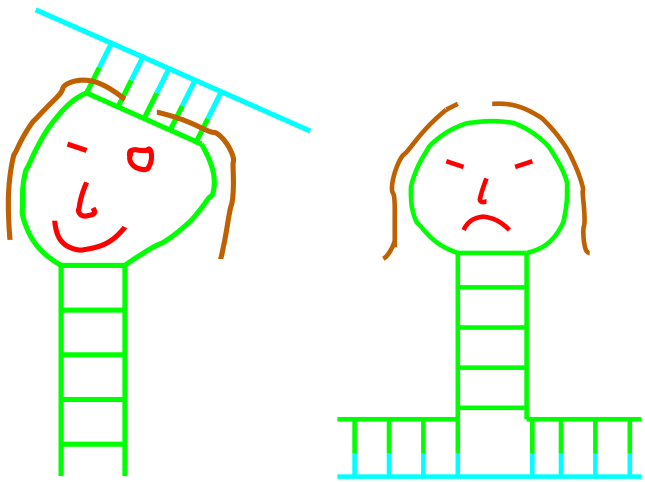
mutant
622-66, 50%



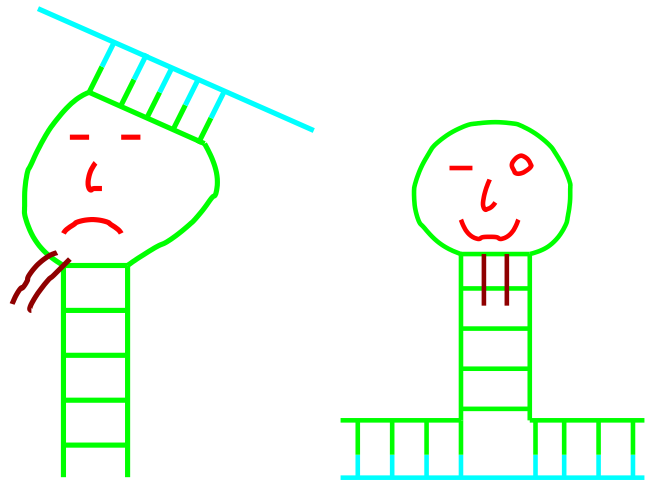
Outline

Problems

RNAup



RNAcofold



Solution?

- Great Unified Cofolding Algorithm
- Cofolding including intermolecular Pseudo-knots and pair probabilities
- Will be expensive computationally and in memory requirements

Unification of RNAcofold and RNAup

First assume no intramolecular base pairs in smaller molecule

- Start with Cofold $\mathcal{O}((n_A + n_B)^3)$ proc. $\mathcal{O}((n_A + n_B)^2)$
- Use RNAup routines, compute intermolecular partition functions ($\mathcal{O}(n^3)$)
- Use RNAup routines, compute partition functions given that pair i, j is innermost base-pair enclosing a binding site. ($\mathcal{O}(n^3)$)

Unification II

- Compute partition functions given duplex between i and j (Like Fold, ($\mathcal{O}(n^3)$))
- Compute partition functions for intra-molecular pairs given Pseudo-knot. $\mathcal{O}(n^3)$
- Run modified RNAup to get the partition functions of intermolecular pairs given Pseudo-knot. $\mathcal{O}(n^3)$

Unification III

- Will take up to six times longer than a simple Cofold
- Will take four to six times more memory.
- Still has time complexity of $\mathcal{O}(n^3)$ in processor time and $\mathcal{O}(n^2)$ in memory.

Thanks to

Ivo Hofacker
Christoph Flamm
Andrea Tanzer
Peter Stadler
...and the wonderful audience