snoRNA-target interactions
part of “interactions of RNAs and proteins”

Sonja Prohaska

Computational EvoDevo
University Leipzig

June 2, 2014
Structure of box H/ACA snoRNAs

Function: guide pseudouridination of uridine by dyskerin

- has two sequence motifs:
  - H box: 5’-ANANNA-3’, ACA box: 5’-ACA-3’
  - target RNA binds to an internal loop of the snoRNA
  - only few mismatches and symmetric internal loops of length 2 or 4
  - target position for pseudouridylation between the two duplexes
  - enzyme: DKC1 = Dyskerin, pseudouridine synthase
  - makes complex with: GAR1, NHP2, NOP10, DKC1
Classes of snoRNAs

names of proteins are taken from yeast
Prediction of H/ACA box snoRNA targets

**RNAsnoop**

- two stem-loop structures
- dynamic programming, thermodynamic folding
- calculate (unbranched) hairpin loop (H)
- calculate left duplex between snoRNA and target RNA (L)
- calculate right duplex between snoRNA and target RNA (R)
- machine learning trained on set of functional snoRNAs

**constraints:**

- take T as given
- duplexes allow only single and tandem mismatches
- target position (i-2) of U to be modified follows left duplex
Prediction of H/ACA box snoRNA targets

- use pattern search and RNA folding to find the 'ACA' and 'ANNANA' motifs and the two stem-loop components
- compute interaction structure separately for the two structures
- calculate left side $L$, calculate unbranched fold $M$ then calculate right side $R$
RNAsnoop: unbranched fold $M$

- consider the *intra*-molecular interaction of subsequence $y[p, q]$ (same as $(y_p...y_q)$) of the snoRNA sequence $y$
- build the best stem-loop structure $M_{p,q}$ from
  - a hairpin loop $\mathcal{H}$
  - and interior loops (incl. stacked basepairs) $\mathcal{I}$
  - if $k = 1$ and $l = 1$ the “interior loop” is a stacked basepair
  - if $k > 1$ or $l > 1$ it is an interior loop
**RNAsnoop: unbranched fold $M$**

\[
M_{p,q} = \min \begin{cases} 
\mathcal{H}(y[p, q]) \\
\min_{k,l} M_{p-k, q+l} + \mathcal{I}(y[p-k, p], y[q, q+l])
\end{cases}
\]  

- either start with the hairpin loop $\mathcal{H}$
- or extend the stem with an interior loop $\mathcal{I}$
RNAsnoop: Left side $L$

- Look at basepair $y_i, x_j$
- $x$ is the target RNA, $y$ is the snoRNA
- Index $i$ runs along the target RNA $x$ (from 5' to 3')
- Index $j$ runs along the target RNA $y$ (from 5' to 3')
- Only symmetric interior loops of length 2 or 4 are allowed in $L$
- To $L_{i-k,j+k}$ we add the an interior loop $I$

$k \neq l$
RNAsnoop: Left side

\[
L_{i,j} = \min_{k=1,2,3} L_{i-k,j+k} + \mathcal{I}(x[i-k,i], y[j,j+k])
\]  

(2)
looking at closing basepair $y_i, x_j$
we combine the left side $L$, the stem-loop $M$ and the pseudouridine-loop
the pseudouridinylation site at $i - 2$ and a nucleotide at $i - 1$ need to be unpaired ($\Psi$-$N$)
$x_{i-3}$ has to contribute a basepair to $L$
x$_i$ has to contribute a basepair to $R$
RNAsnoop: Right side

\[ R_{i,j} = \min \begin{cases} 
\min_{k,l \leq 2} R_{i-k,j+l} + I(x[i-k,i], y[j,j+l]) \\
\min_{l \in [3, |y|-j]} L_{i-3,j+l+1} + M_{j+1,j+l} 
\end{cases} \]  

(3)

- the second term starts the right duplex by adding the left side \( L \) and the stem-loop \( M \)
- the first term continues the right side with an interior loop (or stacked pair) that has the same constraints on loop sizes as the left side
- notice the constraints on the positioning of \( \Psi \) at \( x_{i-2} \)
Tafer H, Kehr S, Hertel J, Hofacker IL and Stadler PF. (2010) 
*RNAsnoop: efficient target prediction for H/ACA snoRNAs*. 
Bioinformatics 26(5), 610-616

Williams GT and Farzaneh F. (2012) *Are snoRNAs and snoRNA host genes new players in cancer?* Nature reviews 12, 84-88