Lecture 8: RNA-RNA interaction

Hamidreza Chitsaz

Colorado State University
chitsaz@cs.colostate.edu

Spring 2015
February 6, 2015
Central dogma

DNA → RNA → Protein
Motivation

Post-transcriptional regulation of gene expression
Regulatory RNA
Repression example (Argaman and Altuvia, J. Mol. Biol. 2000)
Regulatory RNA

Activation example (Repoila, Majdalani, and Gottesman, Mol. Microbiol. 2003)
Background
RNA-RNA MFE structure prediction

- **Avoid intramolecular base pairing**
  - **RNAhybrid** (Rehmsmeier *et al.* 2004), **RNAduplex** (Bernhart *et al.* 2006), **UNAFold** (Markham *et al.* 2008)
  - No internal structure

- **Concatenate input sequences as a single strand; no pseudoknots**
  - **PairFold** (Andronescu *et al.* 2005), **RNACofold** (Bernhart *et al.* 2006)
  - No kissing hairpins

- **Predict binding sites**
  - **RNAup** (Mückstein *et al.* 2008), **intaRNA** (Busch *et al.* 2008)
  - Just one binding site not complete structure

- **Concatenate input sequences; consider special pseudoknots**
  - **NUPACK** (Dirks *et al.* 2003, 2007)
  - Still no kissing hairpins!
Background
RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  - \textit{RNAhybrid} (Rehmsmeier \textit{et al.} 2004), \textit{RNAduplex} (Bernhart \textit{et al.} 2006), \textit{UNAFold} (Markham \textit{et al.} 2008)
  - No internal structure

- Concatenate input sequences as a single strand; no pseudoknots
  - \textit{PairFold} (Andronescu \textit{et al.} 2005), \textit{RNACofold} (Bernhart \textit{et al.} 2006)
  - No kissing hairpins

- Predict binding sites
  - \textit{RNAup} (Mückstein \textit{et al.} 2008), \textit{intaRNA} (Busch \textit{et al.} 2008)
  - Just one binding site not complete structure

- Concatenate input sequences; consider special pseudoknots
  - \textit{NUPACK} (Dirks \textit{et al.} 2003, 2007)
  - Still no kissing hairpins!
Background
RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  RNAhybrid (Rehmsmeier et al. 2004), RNAduplex (Bernhart et al. 2006), UNAFold (Markham et al. 2008)
  No internal structure

- Concatenate input sequences as a single strand; no pseudoknots
  PairFold (Andronescu et al. 2005), RNAcofold (Bernhart et al. 2006)
  No kissing hairpins

- Predict binding sites
  RNAup (Mückstein et al. 2008), intaRNA (Busch et al. 2008)
  Just one binding site not complete structure

- Concatenate input sequences; consider special pseudoknots
  NUPACK (Dirks et al. 2003, 2007)
  Still no kissing hairpins!
Background
RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  
  RNAhybrid (Rehmsmeier et al. 2004), RNAduplex (Bernhart et al. 2006), UNAFold (Markham et al. 2008)
  No internal structure

- Concatenate input sequences as a single strand; no pseudoknots
  
  PairFold (Andronescu et al. 2005), RNAcofold (Bernhart et al. 2006)
  No kissing hairpins

- Predict binding sites
  
  RNAup (Mückstein et al. 2008), intaRNA (Busch et al. 2008)
  Just one binding site not complete structure

- Concatenate input sequences; consider special pseudoknots
  
  NUPACK (Dirks et al. 2003, 2007)
  Still no kissing hairpins!
Consider inter- and intramolecular base pairing

IRIS (Pervouchine 2004), inteRNA (Alkan et al. 2005), Grammatical Approach (Kato et al. 2009)

Voilà, now we are talking business.

The problem is NP-Hard (Alkan et al. 2005); no surprise as pseudoknots are NP-Hard. Exclude zigzags and crossing interactions to lift the curse of complexity and obtain an exact \(O(n^6)\)-time \(O(n^4)\)-space DP algorithm (albeit for simple base-pair counting).

First order zigzag. A general zigzag involves an arbitrary number of kissing hairpins.
Ahhh...but MFE is often wrong!

*Question*: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

*Answer*: the key enabling technology is the **partition function**. All of the above can be computed from the partition function.
Ahhh...but MFE is often wrong!

*Question:* how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

*Answer:* the key enabling technology is the **partition function**. All of the above can be computed from the partition function.
Ahhh...but MFE is often wrong!

Question: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

Answer: the key enabling technology is the partition function. All of the above can be computed from the partition function.
Ahhh...but MFE is often wrong!

*Question:* how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

*Answer:* the key enabling technology is the *partition function*. All of the above can be computed from the partition function.
Partition function

\[ Q(T) = \sum_{f \in F} e^{-G_f/RT}, \]

\[ F = \text{All permissible interaction structures}, \]

\[ p(f) \propto e^{-G_f/RT}, \]

and \( Q \) is the normalizing factor. Also other thermodynamic quantities can be derived from \( Q \).
Partition function

\[ Q(T) = \sum_{f \in F} e^{-G_f/RT}, \]

\( F = \) All permissible interaction structures,

\[ p(f) \propto e^{-G_f/RT}, \]

and \( Q \) is the normalizing factor. Also other thermodynamic quantities can be derived from \( Q \).
Our extension of the Turner model

Chitsaz et al., Bioinformatics 25(12): i365-i373

Hybrid component: as if intramolecular, with penalties.
Kissing loop: like multibranch loop.
Partition function for two strands

**straight vertical line:** intermolecular bond  
**solid:** a base pair  
**dotted:** not a base pair  
**dashed:** either of those two

\[ Q_{i_R,j_R,i_S,j_S} = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{i_R \leq k_1 < j_R, \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{k_1,j_R,i_S,k_2} + \sum_{i_R \leq k_1 < j_R, \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{k_1,i_R,i_S,k_2}. \]
Partition function for two strands

straight vertical line: intermolecular bond
solid: a base pair
dotted: not a base pair
dashed: either of those two

\[
Q_{i_R,j_R,i_S,j_S}^I = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{i_R \leq k_1 < j_R \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{i_R,j_R,i_S,k_2} + \\
\sum_{i_R \leq k_1 < j_R \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{i_R,j_R,i_S,k_2}.
\]
Partition function for two strands

*straight vertical line:* intermolecular bond
*solid:* a base pair
*dotted:* not a base pair
*dashed:* either of those two

\[
Q_{i_R,j_R,i_S,j_S}^{I} = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{i_R \leq k_1 < j_R \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{k_1,j_R,i_S,k_2}^{Ia} + \sum_{i_R \leq k_1 < j_R \atop i_S < k_2 \leq j_S} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q_{k_1,j_R,i_S,k_2}^{Ib}.
\]
$Q^{Ib}$

$b$: stands for bond
\( Q^{Ia} \)

\( a\): stands for arc

\( s\): stands for subsume

\( e\): stands for equivalent

\[
\begin{align*}
{i_R} & \quad j_R \\
I_a & \\
i_S & \quad i_S
\end{align*}
\]

\[
\begin{align*}
I_s & \quad I \\
k_1 & \\
k_2 & \\
I_{s'} & \quad I \\
k_1 & \\
k_2 & \\
I_e & \quad I
\end{align*}
\]
$Q^{Is}$ and $Q^{Le}$

$s$: stands for subsume
$k$: stands for kissing-loop
$m$: stands for multi-loop

e: stands for equivalent
All tables
All tables
All tables
All tables

\[ I_{nn} = I_{sn} \]

\[ I_{dn} \]
All tables

\[ I_b = I_h \]

\[ j_s, i_s \]

\[ k_1 \]

\[ k_2 \]

\[ I_h \]

\[ I_a \]

\[ k_1' \]

\[ k_2' \]

\[ b_z \]

\[ k_1 \]

\[ k_2 \]

\[ I_h \]

\[ I_b \]
All tables

\[ I_{d*} = I_{dn} = I_{dd} \]
All tables

\[ I_{dn} = I_{a_{dn}} \]
All tables
All tables

\[ i_R \quad j_R \quad \quad Ih \quad \quad i_S \quad j_S \quad = \quad Ih \quad \quad k_1 \quad k_2 \]
All tables

\[
\begin{align*}
&\begin{array}{c}
\text{\( I_{n*} \)}\\ \text{\( I_{nn} \)}\\ \text{\( I_{nd} \)}
\end{array}
\end{align*}
\]
All tables
All tables
All tables
All tables
All tables

\[
\begin{align*}
    \begin{tikzpicture}
        \draw (0,0) -- (1,0) node[midway, below] {d} node[midway, above] {e};
        \draw[dashed] (0,0) circle (1cm) node[midway, above] {gk};
        \draw (0.5,0) node[midway, below] {i} -- (1,0) node[midway, above] {j};
    \end{tikzpicture}
    & =
    \begin{tikzpicture}
        \draw (0,0) -- (1,0) node[midway, below] {d} node[midway, above] {e};
        \fill[red] (0.5,0) circle (0.1cm) node[midway, below] {i} -- (1,0) node[midway, above] {j};
        \draw (0,0) circle (1cm) node[midway, above] {gk};
    \end{tikzpicture}
    \\
    \begin{tikzpicture}
        \draw (0,0) -- (1,0) node[midway, below] {d} node[midway, above] {e};
        \fill[red] (0.5,0) circle (0.1cm) node[midway, below] {i} node[midway, above] {k_1} -- (1,0) node[midway, above] {k_2} node[midway, above] {j};
        \draw (0,0) circle (1cm) node[midway, above] {g};
    \end{tikzpicture}
\end{align*}
\]
All tables

\[
\begin{align*}
I_{a_{dn}} &= I_{sm} + I_{sk'} + I_{sm'} + I_e \\
I_{dn} &= \sum_{k_1, k_2} k_1 \cdot k_2 \\
I_{sm} &= \sum_{k_1, k_2} k_1 \cdot k_2 \\
I_{sk'} &= \sum_{k_1, k_2} k_1 \cdot k_2 \\
I_{sm'} &= \sum_{k_1, k_2} k_1 \cdot k_2 \\
I_e &= \sum_{k_1, k_2} k_1 \cdot k_2
\end{align*}
\]
All tables

\[ I_{a_{nd}} = I_{sm} = I_{sk} = I_{sm'} = I_e \]
All tables
All tables

\[ I_{nn} = I_{sm} I_{nn} I_{sm'} I_{mn} I_{en} \]
All tables

=
All tables

\[ \begin{align*}
\begin{array}{c}
I_{b_r} \\
\begin{array}{c}
\begin{array}{c}
i_R \\
\begin{array}{c}
j_R \\
\begin{array}{c}
js \\
\begin{array}{c}
i_S \\
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
= \\
\begin{array}{c}
I_h \\
\begin{array}{c}
\begin{array}{c}
k_1 \\
I_{hb} \\
k_2 \\
I_{b_r} \\
\end{array}
\end{array}
\end{array}
\end{array}
\end{align*} \]
\( \text{All tables} \)

\[
\begin{align*}
I_e &= I_{sm} \\
I_{sk} &= \frac{1}{k_1 - k_2}
\end{align*}
\]
All tables
All tables
All tables

\[ \text{Ihh} \]
All tables

\[
I_{km} = I_{a_{dn}} I_{sm} I_{a_{dn}} I_{s'} I_{a_{dn}} I_{e} I_{a_{nn}} I_{sk'}
\]
All tables
All tables
All tables

\[ I_a = I_r + I_b = I_a \]
Equilibrium concentrations
For two RNAs $R$ and $S$

Assume five types of chemical compounds: $R$, $S$, $RR$, $SS$, $RS$.

Solve

\[
K_R = \frac{Q_{RR}^I}{Q_R^2} = \frac{N'_{RR}}{N_R^2},
\]

\[
K_S = \frac{Q_{SS}^I}{Q_S^2} = \frac{N'_{SS}}{N_S^2},
\]

\[
K_{RS} = \frac{Q_{RS}^I}{Q_R Q_S} = \frac{N'_{RS}}{N_R N_S},
\]

\[
N_{RS} = N_R^0 - 2N_{RR} - N_R = N_S^0 - 2N_{SS} - N_S,
\]

to obtain the equilibrium concentrations $N$. $N^0$ are the initial concentrations of single strands.
Equilibrium concentration of OxyS with wild type fhIA

Init. $[\text{OxyS}] = 2\text{nM}$, $[\text{fhIA}] = 0 \text{ to } 1000\text{nM}$
Equilibrium concentration of OxyS with fhlA mutants

- $fhl_{A8G}$
- $fhl_{C13G}$
- $fhl_{G37C;G38C}$
- $fhl_{G38C;G39C}$

Our Algorithm vs. Experiment
## Melting temperature prediction
Comparison of piRNA results over three data sets

<table>
<thead>
<tr>
<th>Set</th>
<th>Size</th>
<th>Length</th>
<th>piRNA Avg error</th>
<th>RNAcofold Avg error</th>
<th>UNAFold Avg error</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>9 short pairs</td>
<td>5-7nt</td>
<td>1.48°C</td>
<td>9.35°C</td>
<td>8.55°C</td>
</tr>
<tr>
<td>II</td>
<td>12 pairs</td>
<td>∼ 20nt</td>
<td>4.86°C</td>
<td>22.97°C</td>
<td>9.12°C</td>
</tr>
<tr>
<td>III</td>
<td>62 pairs</td>
<td>22 – 40nt</td>
<td>1.91°C</td>
<td>14.34°C</td>
<td>26.53°C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>Size</th>
<th>Length</th>
<th>Spearman rank correlation</th>
<th>piRNA</th>
<th>RNAcofold</th>
<th>UNAFold</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>9 short pairs</td>
<td>5-7nt</td>
<td>0.97</td>
<td>0.97</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>12 pairs</td>
<td>∼ 20nt</td>
<td>0.41</td>
<td>-0.03</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>62 pairs</td>
<td>22 – 40nt</td>
<td>0.3</td>
<td>-0.04</td>
<td>0.24</td>
<td></td>
</tr>
</tbody>
</table>