Incorporating RNA-Protein Interactions into RNA Secondary Structure Prediction

Ralf Bundschuh

Departments of Physics and Chemistry&Biochemistry, Center for RNA Biology
Interdisciplinary Biophysics Graduate Program, Ohio State University

July 29, 2012

Overview

1. Introduction
2. Physical extension
3. Interaction with proteins
4. Putting it all together
5. Summary

Overview

1. Introduction
2. Physical extension
3. Interaction with proteins
4. Putting it all together
5. Summary

RNA secondary structure prediction

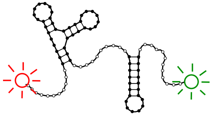
Secondary structure prediction

Prediction of RNA secondary structures in solution is mature field: Nussinov, Waterman, McCaskill, MFOLD, Dynafold, Vienna package, Sfold, . . .

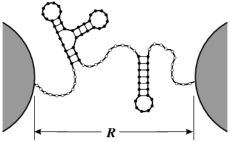
But . . .

Solution structure may not always be what we want ⇒

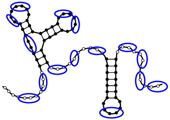
FRET



Force-extension



Protein-RNA interactions



Partition function approach

Partition function approach (McCaskill, 1990)

- Each secondary structure S is a state
- Each state S has a (free) energy

$$E[S] = \sum_{(i,j)\in S} \varepsilon_{i,j}$$

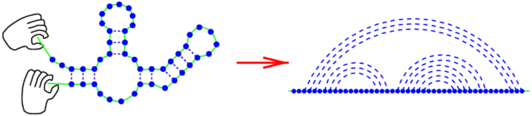
(in practice use much more detailed energy model)

- All thermodynamic quantities can be derived from partition function

$$Z = \sum_{S} \exp\left(-E[S]/k_B T\right)$$

Arc diagrams

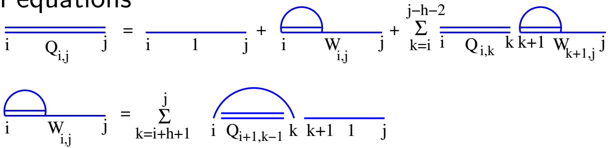
Secondary structures correspond one to one to arc diagrams



Partition functions as recursion equations

- Partition function $Q_{i,j}$  for substrand from base *i* to base *j*
- Partition function $W_{i,j}$  for substrand from base *i* to base *j* where base *i* is paired to the last paired base

⇒ Recursion equations

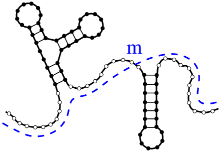


Overview

1. Introduction
2. Physical extension
    - Restricted partition function
    - Polymer Physics
    - FRET
3. Interaction with proteins
4. Putting it all together
5. Summary

External bases

Physical extension of a given secondary structure depends only on the external bases m.



Restricted partition functions

Define $Q(m)$ to be the partition function over all structures with fixed number of external bases m.

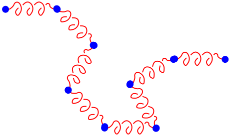
- Can be calulated by similar recursion equation for $Q_{1,j}(m)$
- $Z = \sum_{m=\Delta}^{N} Q(m)$
- $P(m) = Q(m)/Z$ is the probability to observe m external bases

U. Gerland, RB, and T. Hwa, Biophys J., 2001

Polymer Physics

Backbone model

- For given m backbone is a single-stranded RNA molecule
- Well described by extensible freely jointed chain
- Distribution $W(R|m)$ of end-to-end-distance R well known



Including secondary structure

End-to-end distance distribution

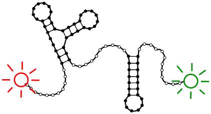
$$P(R) = \sum_{m=\Delta}^{N} P(m) W(R|m)$$

Force-distance relationship

$$F(R) = -k_B T \frac{\partial}{\partial R} \log \sum_{m=\Delta}^{N} Q(m) W(R|m)$$

FRET

Problem



- Energy transfer proportional to $(R_0/R)^6$
- Backbone moves faster than FRET time scale

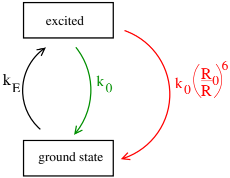
$\Rightarrow$ Energy transfer is not given by $(R_0/\langle R\rangle)^6$

Solution

- Monte-Carlo simulation of polymer dynamics

  Murphy et al., Biophys. J. 2004
- Provides $P(E_{FRET}|m)$
- Total expected FRET-distribution

$$P(E_{FRET}) = \sum_{m=\Delta}^{N} P(E_{FRET}|m)P(m)$$



excited

ground state

k_E

k_0

$k_0\left(\frac{R_0}{R}\right)^6$

Overview

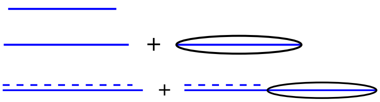
1. Introduction
2. Physical extension
3. Interaction with proteins
    - Protein on single-stranded RNA
    - Including secondary structure
4. Putting it all together
5. Summary

Protein on single-stranded RNA

Main quantity

Define $O_{i,j}$ to be the partition function of the single-stranded RNA from base i to base j in the presence of a single-stranded binding protein at concentration C with footprint F and dissociation constant K_D.

Calculation

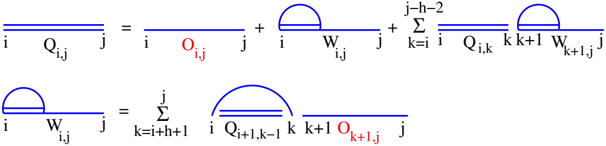


- $O_{i,j} = 1$ if $j - i + 1 < F$
- $O_{i,i+F-1} = 1 + C/K_D$
- $O_{i,j} = O_{i,j-1} + C/K_D O_{i,j-F}$
- Can include sequence dependence and cooperativity

RNA secondary structure with protein interactions

Secondary structure recursion

Protein-RNA interaction can be incorporated into secondary structure recursions by replacing 1 by $O_{i,j}$



Overview

1. Introduction
2. Physical extension
3. Interaction with proteins
4. Putting it all together
    - FRET prediction
    - Base pairing and protein binding
5. Summary

FRET prediction

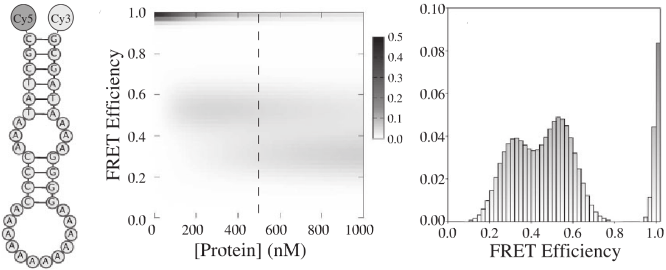
Implementation

- Incorporate everything into Vienna package

  Hofacker et al., Monatsh. Chemie, 1994
- Uses full nearest neighbor model and Turner parameters

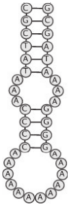
Example

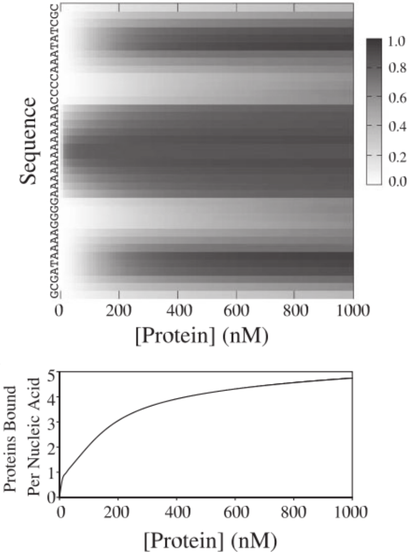
Protein with binding energy of 10kcal/mol at 1M and 6nt footprint



Other observables

- Can extend ability to calculate pair probabilities to case with proteins
- Can calculate protein occupancy of every base
- Can calculate number of bound proteins by taking numerical derivative





Overview

1. Introduction
2. Physical extension
3. Interaction with proteins
4. Putting it all together
5. Summary

Summary

Conclusions

- Physical end-to-end distance can be incorporated into RNA secondary structure prediction
- Protein-RNA interactions can be included in RNA secondary structure

Future work

- Apply approach to extract mechanistic parameters from experiments
- Double-stranded RNA-binding proteins
- RNA kinetics in the presence of proteins

Acknowledgements

Bob Forties

NSF DMR-0706002