Statistical Physics of RNA folding

Ralf Bundschuh

Collaborators: Terry Hwa, UCSD, Ulrich Gerland, Universität zu Köln, Robijn Bruinsma, UCLA

Ohio State University

March 18, 2008
Outline of all lectures

- Part I: Introduction to RNA biology
- Part II: Homogeneous RNA
- Part III: Sequence disorder
- Part IV: Force-extension experiments
Part I

Introduction to RNA biology
Outline of part I

1. RNA structure
2. Making RNA
3. Functions of RNA
4. Summary
Outline of part I

1. RNA structure
   - Chemical structure
   - Secondary structure
   - Tertiary structure

2. Making RNA

3. Functions of RNA

4. Summary
Chemical structure of RNA

- RNA is a heteropolymer
- Backbone is phosphate-sugar chain
  - Directionality
  - Quite flexible
  - Negatively charged in solution
- Four different side chains ("bases")
  - Hydrophobic
  - Two groups: purines (A,G) and pyrimidines (C,U)
Secondary structure of RNA

- **Primary structure**: sequence (5’ → 3’)
  
  GCGGAUUUAGCUAGUUGGAGAGCGCCAGACUGAAAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA

- **Secondary structure**: base pairing
  - Watson-Crick (G–C and A–U)
  - Wobble (G–U)
  - non-canonical (anything else)
  - Allows formation of hydrogen bonds
  - Protects hydrophobic bases from water
  - Costs configurational entropy
  - Brings like charges close together
  - Free energy per (WC) base pair $\approx 2 \frac{kcal}{mol} \approx 4k_B T$
  - Strongly dependent on temperature and salt concentration
Tertiary structure of RNA

- **Tertiary structure**: three-dimensional arrangement
  - Weaker interactions than base pairing
  - Depends on Magnesium
  - Confers biological function
Outline of part I

1 RNA structure
2 Making RNA
3 Functions of RNA
4 Summary
Template for RNA is stored in genome
RNA is produced during transcription
RNA is produced by RNA polymerase
RNA is exact copy of DNA sequence

Transcription_(DNA_transcription)
Outline of part I

1. RNA structure
2. Making RNA
3. Functions of RNA
   - Messenger RNA
   - Transfer RNA
   - Ribosomal RNA
   - Other RNA
4. Summary
Translation

- **Messenger RNA** codes for protein sequence
- **Translation** from messenger RNA occurs in **ribosome**
- Three bases (a **codon**) in the messenger RNA code for one amino acid via the **genetic code**
- RNA structure is **not very important**
Transfer RNA

- Adapter molecule that implements genetic code
- Binds amino acid at one end
- Base pairs with messenger RNA at other end
- Very specific clover leaf structure recognized by ribosome

wikipedia
http://ghs.gresham.k12.or.us/science/ps/sci/ibbio/chem/nucleic/chpt15/tRNA.gif
Ribosomal RNA

- **Ribosomal RNA** is major component of the ribosome
- **Ribosomal RNA** catalyzes actual extension of protein sequence
- Overall structure **conserved** throughout all of cellular life

RNA world hypothesis
Other RNA

- **snoRNA** guides chemical modification of other RNAs
- **microRNA** regulates the translation of other genes
- **Ribozymes** catalyze chemical reactions
  (RNaseP, group I and II intron, Hammerhead ribozyme, . . .)
- **Guide RNA** guides RNA editing
- **RNA viruses** use single or double stranded RNA as their genome
- **New categories** found all the time . . .
  (ENCODE project finds that very large fraction of human genome is transcribed, but only few percent are protein coding)
Outline of part I

1. RNA structure
2. Making RNA
3. Functions of RNA
4. Summary
Summary of part I

- RNA is a heteropolymer of four bases
- RNA forms a well-defined hierarchy of structures
- RNA is involved in nearly every aspect of life
Part II

Homogeneous RNA
Outline of part II

5 Boltzmann partition function

6 Molten RNA

7 Molten-native transition — hairpin

8 Molten-native transition — Cayley tree

9 Summary
Outline of part II

5 Boltzmann partition function
   - Secondary structure
   - Partition function
   - Recursion equation

6 Molten RNA

7 Molten-native transition — hairpin

8 Molten-native transition — Cayley tree

9 Summary
Definition of RNA secondary structure

Definition

An RNA secondary structure on an RNA sequence of length $N$ is a set $S$ of base pairs $(i, j)$ with $1 \leq i < j \leq N$ that fulfill the following conditions:

- Each base is involved in at most one base pair
- No pseudoknots, i.e., if $(i, j)$ and $(k, l)$ are base pairs with $i < k$, either $i < k < l < j$ or $i < j < k < l$
Diagrammatic representation

An RNA secondary structure can be represented by an arch diagram.

- Every base is end point to at most one arch
- Two arches never cross
- One to one correspondence between arch diagrams that fulfil the above conditions and secondary structures
Energy models

- Each structure $S$ has a certain (free) energy $E[S]$
- Different models possible
  - All base pairs are equal $E[S] = \varepsilon_0 |S|$ ($\varepsilon_0 < 0$)
  - Energy associated with base pairs $E[S] = \sum_{(i,j)\in S} e(i,j)$
  - Turner parameters (nearest neighbor model): energy associated with loops
    - Base pair stacks
    - Hairpin loops
    - Interior loops
    - Bulges
    - Multi-loops

![Diagram of RNA structures]
Partition function

Definition
The partition function of an RNA molecule with energy function $E[S]$ is given by

$$Z = \sum_S e^{-\frac{E[S]}{RT}}$$

- $R = 1.987 \frac{\text{cal}}{\text{mol} \cdot \text{K}}$
- Temperature $T$ in Kelvin $\Rightarrow RT \approx 0.6 \frac{\text{kcal}}{\text{mol}}$
- Ensemble free energy $F = -RT \ln Z$
- Thermodynamics completely specified if partition function is known
Calculating partition functions

**Definition**
Let $Z_{i,j}$ be the partition function for the RNA molecule starting at base $i$ and ending at base $j$. Denote this quantity by $Z_{i,j}$.

**Observation**
For the base pairing energy model the $Z_{i,j}$ obey the recursion equation

$$Z_{i,j} = Z_{i,j-1} + \sum_{k=i}^{j-1} Z_{i,k-1} e^{-\frac{e(k,j)}{RT}} Z_{k+1,j-1}$$

- Calculate from shortest to longest substrands
- $O(N^3)$ algorithm for arbitrary sequence
Outline of part II

5 Boltzmann partition function

6 Molten RNA
   • Energy model
   • z transform
   • Properties
   • Geometric derivation

7 Molten-native transition — hairpin

8 Molten-native transition — Cayley tree

9 Summary
Energetics in molten phase

Definition

In the molten phase of RNA every base can pair with every other base equally well, i.e., \( e(i, j) = \varepsilon_0 \).

Properties

- Applies to repetitive sequences: AUAUAUAUAUAUAUAUAU, GCGCGCGCGCGCGC, GACGACGACGACGACGACGACGAC
- Applies to arbitrary sequences at temperatures close to denaturation on a coarse-grained level
- Minimum energy is always \( \frac{N}{2} \varepsilon_0 \)
- Structural entropy plays a major role
Reminder

\[ Z_{i,j} = Z_{i,j-1} + \sum_{k=i}^{j-1} Z_{i,k-1} e^{-\frac{e(k,j)}{RT}} Z_{k+1,j-1} \]

Simplification

In the molten phase the partition function depends only on the length \( j - i \) of the substrand, not on \( i \) and \( j \) individually: \( Z_{i,j} \equiv G(j - i + 2) \)

Consequence

\[ G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k) G(N - k) \quad \text{with} \quad q \equiv e^{-\frac{\epsilon_0}{RT}} \]
**z transform**

**Definition**

For any series $Q(N)$ the z transform $\hat{Q}$ is defined as

$$\hat{Q}(z) \equiv \sum_{N=1}^{\infty} Q(N) z^{-N}$$

**Properties**

- Function of the complex variable $z$
- Analytic outside a radius of convergence
- Discrete version of Fourier transform
- Back transform: $Q(N) = \frac{1}{2\pi i} \oint \hat{Q}(z) z^{N-1} dz$
- Convolution property: $\sum_{k=1}^{N-1} Q(k) W(N - k) = \hat{Q}(z) \cdot \hat{W}(z)$
$\hat{G}(z)$ can be calculated:

\[
\hat{G}(z) = \frac{1}{2q}\left[ (z - 1) - \sqrt{(z - 1)^2 - 4q} \right]
\]
\( \hat{G}(z) \) can be calculated:

\[
G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k)G(N - k)
\]
$\hat{G}(z)$ can be calculated:

$$G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k) G(N - k)$$

$$G(N + 1)z^{-N} = G(N)z^{-N} + qz^{-N} \sum_{k=1}^{N-1} G(k) G(N - k)$$
\( \hat{G}(z) \) can be calculated:

\[
G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k) G(N - k)
\]

\[
G(N + 1)z^{-N} = G(N)z^{-N} + qz^{-N} \sum_{k=1}^{N-1} G(k) G(N - k)
\]

\[
\sum_{N=1}^{\infty} G(N + 1)z^{-N} = \hat{G}(z) + q\hat{G}^2(z)
\]
\( \hat{G}(z) \) can be calculated:

\[
G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k) G(N - k)
\]

\[
G(N + 1) z^{-N} = G(N) z^{-N} + q z^{-N} \sum_{k=1}^{N-1} G(k) G(N - k)
\]

\[
\sum_{N=1}^{\infty} G(N + 1) z^{-N} = \hat{G}(z) + q \hat{G}^2(z)
\]

\[
z \hat{G}(z) - G(1) = \hat{G}(z) + q \hat{G}^2(z)
\]
$\hat{G}(z)$ can be calculated:

\[
G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k)G(N - k)
\]

\[
G(N + 1)z^{-N} = G(N)z^{-N} + qz^{-N} \sum_{k=1}^{N-1} G(k)G(N - k)
\]

\[
\sum_{N=1}^{\infty} G(N + 1)z^{-N} = \hat{G}(z) + q\hat{G}^2(z)
\]

\[
z\hat{G}(z) - G(1) = \hat{G}(z) + q\hat{G}^2(z)
\]

\[
z\hat{G}(z) - 1 = \hat{G}(z) + q\hat{G}^2(z)
\]
z transform for molten RNA

\( \hat{G}(z) \) can be calculated:

\[
\hat{G}(z) = \frac{1}{2q} \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right]
\]

\[
G(N + 1) = G(N) + q \sum_{k=1}^{N-1} G(k)G(N - k)
\]

\[
G(N + 1)z^{-N} = G(N)z^{-N} + qz^{-N} \sum_{k=1}^{N-1} G(k)G(N - k)
\]

\[
\sum_{N=1}^{\infty} G(N + 1)z^{-N} = \hat{G}(z) + q\hat{G}^2(z)
\]

\[
z\hat{G}(z) - G(1) = \hat{G}(z) + q\hat{G}^2(z)
\]

\[
z\hat{G}(z) - 1 = \hat{G}(z) + q\hat{G}^2(z)
\]
Back transform I

Integral expression for $G(N)$

$$G(N) = \frac{1}{2\pi i} \oint \hat{G}(z) z^{N-1} dz$$
Integral expression for $G(N)$

$$G(N) = \frac{1}{2\pi i} \oint \hat{G}(z) z^{N-1} dz$$

$$= \frac{1}{4\pi qi} \oint \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right] z^{N-1} dz$$
Integral expression for $G(N)$

\[ G(N) = \frac{1}{2\pi i} \oint \hat{G}(z) z^{N-1} \, dz \]

\[ = \frac{1}{4\pi q i} \oint \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right] z^{N-1} \, dz \]

\[ = -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q} z^{N-1} \, dz \]
Back transform I

Integral expression for $G(N)$

$$G(N) = \frac{1}{2\pi i} \oint \hat{G}(z) z^{N-1} dz$$

$$= \frac{1}{4\pi q i} \oint \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right] z^{N-1} dz$$

$$= -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q} z^{N-1} dz$$

Singularity structure

- Branch cut from $1 - 2\sqrt{q}$ to $z_0 \equiv 1 + 2\sqrt{q}$
Back transform I

Integral expression for $G(N)$

$$G(N) = \frac{1}{2\pi i} \oint \hat{G}(z) z^{N-1} dz$$

$$= \frac{1}{4\pi qi} \oint \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right] z^{N-1} dz$$

$$= -\frac{1}{4\pi qi} \oint \sqrt{(z - 1)^2 - 4q} z^{N-1} dz$$

Singularity structure

- Branch cut from $1 - 2\sqrt{q}$ to $z_0 \equiv 1 + 2\sqrt{q}$
- Contour has to go around branch cut
Back transform I

Integral expression for $G(N)$

$$G(N) = \frac{1}{2\pi i} \oint \hat{G}(z)z^{N-1}dz$$

$$= \frac{1}{4\pi q i} \oint \left[ z - 1 - \sqrt{(z - 1)^2 - 4q} \right] z^{N-1}dz$$

$$= -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q} z^{N-1}dz$$

Singularity structure

- Branch cut from $1 - 2\sqrt{q}$ to $z_0 \equiv 1 + 2\sqrt{q}$
- Contour has to go around branch cut
- Contour can be contracted to branch cut
Reminder

\[
G(N) = -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q z^{N-1}} \, dz
\]
Reminder

\[ G(N) = -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q z^{N-1}} \, dz \]

Leading behavior

For large \(N\) only the singularity with **largest real part** contributes

\[ G(N) \approx z_0^N = (1 + 2\sqrt{q})^N \]
Reminder

\[ G(N) = -\frac{1}{4\pi q i} \oint \sqrt{(z - 1)^2 - 4q z^{N-1}} \, dz \]

Leading behavior

For large \( N \) only the singularity with largest real part contributes

\[ G(N) \approx z_0^N = \left(1 + 2\sqrt{\bar{q}}\right)^N \]

Prefactor

\[ \int_{\mu_0}^{\mu_c} (\mu_c - \mu)^\alpha e^{\mu N} \, d\mu \approx \Gamma(1 + \alpha) N^{-(1 + \alpha)} e^{\mu_c N} \]

\[ \Rightarrow G(N) \sim N^{-3/2} (1 + 2\sqrt{\bar{q}})^N \]
Properties of molten RNA

\[ G(N) \approx \left( \frac{1 + 2\sqrt{q}}{4\pi q^{3/2}} \right)^{1/2} N^{-3/2} (1 + 2\sqrt{q})^N \]

- \( N^{-3/2} \) characteristic behavior due to entropy
- Can be observed in pairing probability:

\[ \Pr\{1 \text{ and } k \text{ paired}\} = q \frac{G(k)G(N-k)}{G(N+1)} \sim k^{-3/2} \frac{(N-k)^{-3/2}}{N^{-3/2}} \approx k^{-3/2} \]

- Free energy is \( F = -RT \ln G(N) \approx -RTN \ln(1 + 2\sqrt{q}) \)
- For \( q \gg 1 \): \( F \approx -RTN \ln(2\sqrt{q}) = -\frac{RT}{2} N \ln(q) = N \frac{\varepsilon_0}{2} \)
- For \( q = 1 \): \( G(N) \approx \text{number of secondary structures} \sim N^{-3/2} 3^N \).
A geometric argument

Mapping

Properties

- One to one mapping between structures and mountains
- Mountains are random walks
- At each point two choices with weight $\sqrt{q}$ and one with weight 1
  \[ \Rightarrow (1 + 2\sqrt{q})^N \]
- Never go below zero but return at the end \[ \Rightarrow N^{-3/2} \]
Outline of part II

5 Boltzmann partition function

6 Molten RNA

7 Molten-native transition — hairpin
   • Model
   • Solution
   • Results

8 Molten-native transition — Cayley tree

9 Summary
Model for molten-native transition

Observation

Structural RNAs have to fold into a specific "native" structure ⇒ there must be something in the sequence that prefers this structure

Model

- Use perfect hairpin as native structure
  \[
  \begin{array}{c}
  1 \\
  \hline \\
  N/2 \\
  \hline \\
  N
  \end{array}
  \]
- Assign binding energy $\varepsilon_1$ to all native base pairs
- Assign binding energy $\varepsilon_0$ to all other base pairs
Partition function

Definition

Let $Z(N; q, \tilde{q})$ be the partition function of the Gō model for the molten-native transition with $2N - 2$ bases, $q = e^{-\varepsilon_0/RT}$, and $\tilde{q} = e^{-\varepsilon_1/RT}$.

Definition

Let $W(N; q) \equiv Z(N; q, \tilde{q} = 0)$ be the partition function of the Gō model for $2N - 2$ bases in which native contacts are disallowed.

Observation

$Z(N; q, \tilde{q}) = (\text{p.f. with 0 native contacts}) + (\text{p.f. with 1 native contacts}) + (\text{p.f. with 2 native contacts}) + \ldots$
Partition function

Definition
Let \( Z(N; q, \tilde{q}) \) be the partition function of the G\( \ddot{\text{o}} \) model for the molten-native transition with \( 2N - 2 \) bases, \( q = e^{-\varepsilon_0/RT} \), and \( \tilde{q} = e^{-\varepsilon_1/RT} \).

Definition
Let \( W(N; q) \equiv Z(N; q, \tilde{q} = 0) \) be the partition function of the G\( \ddot{\text{o}} \) model for \( 2N - 2 \) bases in which native contacts are disallowed.

Observation
\[ Z(N; q, \tilde{q}) = (\text{p.f. with 0 native contacts}) + (\text{p.f. with 1 native contacts}) + (\text{p.f. with 2 native contacts}) + \ldots \]
\[ = \text{Diagram 1} + \text{Diagram 2} + \ldots + \text{Diagram n} + \ldots \]
Molten-native transition — hairpin

Solution

Individual terms

\[
\begin{align*}
\text{0 native contacts} & : p.f. \text{ with 0 native contacts} = W(N; q) \\
\rightarrow & : z\text{-transform: } \hat{W}(z) \\
\text{1 native contact} & : p.f. \text{ with 1 native contacts} = \tilde{q} \sum_{N-1}^{N} W(k; q) W(N-k; q) \\
\rightarrow & : z\text{-transform: } \tilde{q} \hat{W}^2(z) \\
\text{2 native contacts} & : p.f. \text{ with 2 native contacts} = \tilde{q}^2 \sum_{1 \leq k_1 < k_2 < N} W(k_1; q) W(k_2-k_1; q) W(N-k_2; q) \\
\rightarrow & : z\text{-transform: } \tilde{q}^2 \hat{W}^3(z)
\end{align*}
\]
Individual terms

0 native contacts

(p.f. with 0 native contacts) = \( \mathcal{W}(N; q) \)

⇒ z-transform: \( \hat{\mathcal{W}}(z) \)
Individual terms

0 native contacts

(p.f. with 0 native contacts) = \[ W(N; q) \]
\[ \Rightarrow z\text{-transform: } \hat{W}(z) \]

1 native contact

(p.f. with 1 native contacts) = \[ \tilde{q} \sum_{k=1}^{N-1} W(k; q) W(N - k; q) \]
\[ \Rightarrow z\text{-transform: } \tilde{q} \hat{W}^2(z) \]
Individual terms

0 native contacts

\[(p.f. \text{ with } 0 \text{ native contacts}) = \mathcal{W}(N; q) \Rightarrow z\text{-transform: } \hat{\mathcal{W}}(z)\]

1 native contact

\[(p.f. \text{ with } 1 \text{ native contacts}) = \tilde{q} \sum_{k=1}^{N-1} \mathcal{W}(k; q) \mathcal{W}(N - k; q) \Rightarrow z\text{-transform: } \tilde{q}\hat{\mathcal{W}}^2(z)\]

2 native contacts

\[(p.f. \text{ with } 2 \text{ native contacts}) = \tilde{q}^2 \sum_{1 \leq k_1 < k_2 < N} \mathcal{W}(k_1; q) \mathcal{W}(k_2 - k_1; q) \mathcal{W}(N - k_2; q) \Rightarrow z\text{-transform: } \tilde{q}^2\hat{\mathcal{W}}^3(z)\]
Putting it all together

Summing up

\[ Z(z; \tilde{q}, q) = W(z) + \tilde{q} W_2(z) + \tilde{q}^2 W_3(z) + \ldots \]

What is \( \hat{W}(z) \)?

For \( \tilde{q} = q \) we have

\[ Z(N; q, \tilde{q} = q) = G(2N-1; q) \]

\[ \hat{Z}(z; q, q) = \hat{E}(z; q) \equiv \sum_{N=1}^{\infty} G(2N-1; q) z^{-N} \]

\[ \hat{W}(z) = \hat{E}(z) + q \hat{E}(z) \]

\[ \hat{Z}(z; q, \tilde{q}) = \hat{E}(z) - (\tilde{q} - q) \hat{E}(z) \]
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \]

\[ = \hat{E}(z) + q\hat{E}(z) + \tilde{q}\hat{E}(z) + \tilde{q}^2\hat{E}(z) + \ldots = \]

\[ = \hat{E}(z) + (q + \tilde{q} - q)\hat{E}(z) = \hat{E}(z) \]

What is \( \hat{W}(z) \)?
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)} \]
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)} \]

What is \( \hat{W}(z) \)?
Putting it all together

Summing up

\[
\hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)}
\]

What is \( \hat{W}(z) \)?

For \( \tilde{q} = q \) we have \( Z(N; q, \tilde{q} = q) = G(2N - 1; q) \)
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)} \]

What is \( \hat{W}(z) \)?

For \( \tilde{q} = q \) we have \( Z(N; q, \tilde{q} = q) = G(2N - 1; q) \)

\[ \Rightarrow \hat{Z}(z; q, q) = \hat{E}(z; q) = \sum_{N=1}^{\infty} G(2N - 1; q)z^{-N} \]
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)} \]

What is \( \hat{W}(z) \)?

For \( \tilde{q} = q \) we have
\[
\hat{Z}(z; q, q) = \hat{E}(z; q) \equiv \sum_{N=1}^{\infty} G(2N - 1; q)z^{-N}
\]

\[ \Rightarrow \hat{W}(z) = \frac{\hat{E}(z)}{1 + q\hat{E}(z)} \]
Putting it all together

Summing up

\[ \hat{Z}(z; \tilde{q}, q) = \hat{W}(z) + \tilde{q}\hat{W}^2(z) + \tilde{q}^2\hat{W}^3(z) + \ldots = \frac{\hat{W}(z)}{1 - \tilde{q}\hat{W}(z)} \]

What is \( \hat{W}(z) \)?

For \( \tilde{q} = q \) we have \( Z(N; q, \tilde{q} = q) = G(2N - 1; q) \)

\[ \Rightarrow \hat{Z}(z; q, q) = \hat{E}(z; q) \equiv \sum_{N=1}^{\infty} G(2N - 1; q)z^{-N} \]

\[ \Rightarrow \hat{W}(z) = \frac{\hat{E}(z)}{1 + q\hat{E}(z)} \]

\[ \Rightarrow \hat{Z}(z; q, \tilde{q}) = \frac{\hat{E}(z)}{1 - (\tilde{q} - q)\hat{E}(z)} \]
Behavior of $\hat{E}(z)$

Expression for $\hat{E}(z)$

$$\hat{E}(z) = \sum_{N=1}^{\infty} G(2N - 1; q) z^{-N}$$

can be calculated similarly to $\hat{G}(z)$.

$$\hat{E}(z) = \frac{1}{2q} - \frac{1}{4qz} \left[ \sqrt{(z - 1)^2 - 4q} + \sqrt{(z + 1)^2 - 4q} \right].$$

Properties

- Vanishes as $z \to \infty$
- Square root branch cut at $z_0 = 1 + 2\sqrt{q}$
- Finite limit $\hat{E}(1 + 2\sqrt{q})$ at branch cut
- Other branch cuts have smaller real part
Singularity structure of $\hat{Z}(z)$

Candidate singularities

$$\hat{Z}(z; q, \tilde{q}) = \frac{\hat{E}(z)}{1 - (\tilde{q} - q)\hat{E}(z)}$$

- Square root branch cut at $z_0 = 1 + 2\sqrt{q}$
- Pole at $z_1(\tilde{q})$ given by $\hat{E}(z_1) = 1/(\tilde{q} - q)$

Dominant singularity

- Square root branch cut at $z_0$ if $1/(\tilde{q} - q) > \hat{E}(z_0)$
- Pole at $z_1(\tilde{q})$ if $1/(\tilde{q} - q) < \hat{E}(z_0)$

$\Rightarrow$ Phase transition
Properties of molten-native transition

Characterization of phase transition

- Critical bias $\tilde{q}_c = q + 1/\hat{E}(z_0)$
- If $\tilde{q} < \tilde{q}_c$ regular molten behavior $Z \sim N^{-3/2}(1 + 2\sqrt{\tilde{q}})^N$
  $\longrightarrow$ native base pairs do not play any role
- If $\tilde{q} > \tilde{q}_c$ native behavior with $Z \sim N^0 z_1(\tilde{q})^N$
  $\longrightarrow$ finite fraction of native base pairs but still many “bubbles”
- For $\tilde{q} \gg \tilde{q}_c$ we get $Z \sim N^0 \tilde{q}^N$
  $\longrightarrow$ only native base pairs

Conclusion

It takes a finite amount of sequence bias to enforce a native structure.
Outline of part II

5 Boltzmann partition function

6 Molten RNA

7 Molten-native transition — hairpin

8 Molten-native transition — Cayley tree
   • Model
   • Recursion equations
   • Numerics
   • Analytical solution
   • Open problems

9 Summary
Modeling branchy native states

Problem
Natural RNA structures are branched

Solution
Use Cayley tree as native structure

⇒ Characterize by order $k$ instead of length
$N = 2^{k+2} - 2$
Restricted partition function

Define $Q(k, n)$ as the partition function of all structures with $2n$ bases “outside” the native base pairs and all these bases unpaired.

Full partition function

$W(n)$ is still partition function of non-native base pairs among $2n$ bases.

$\Rightarrow Z(k; q, \tilde{q}) = \sum_n W(n)Q(k, n)$
Recursion equations from level $k$ to $k + 1$ 

$n > 0$

First native base pair is open
⇒ both sub-trees contribute to accessible base pairs
⇒ $Q(k + 1, n) = \sum_m Q(k, m)Q(k, n - m)$

$n = 0$

First native base pair is paired
⇒ both sub-trees contribute to first bubble of non-native base pairs
⇒ $Q(k+1, n=0) = \tilde{q}\sum_{n_1,n_2} Q(k, n_1)Q(k, n_2)W(n_1+n_2)$
Pinch free energy

Free energy

\[ Q(k, n) \longrightarrow Z(k; q, \tilde{q}) \longrightarrow F(k; q, \tilde{q}) = -k_B T \log Z(k; q, \tilde{q}) \]

Expectation

Small \( \tilde{q} \) ⇒ molten phase ⇒ \( F \approx N^{-3/2} z_0^N \)
Large \( \tilde{q} \) ⇒ native phase ⇒ \( F \approx z_1^N \)
\( (N = 2^{k+2} - 2) \)

Pinch free energy

\[ F(k+1) - 2F(k) = \begin{cases} \frac{3}{2} k & \text{molten} \\ \text{const.} & \text{native} \end{cases} \]
Numerically iterate recursion equations for $Q(k, n)$ at $q = 4$.

Phase transition at $\tilde{q} \geq 80$.
Stationary solution

Asymptotics

In native phase: $Z(k) \approx z_1^N$ with $N = 2^{k+2} - 2$

$\Rightarrow$ assume $Q(k, n)/z_1^{2^{k+2}-2} \rightarrow w(n)$ for large $k$

Recursions for $w(n)$

\[
\begin{align*}
w(n) &= \sum_{m=0}^{n-1} w(m)w(n-1-m) \quad \text{and} \\
w(0) &= (\tilde{q} - q) \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} w(n_1)w(n_2)G(2(n_1 + n_2); q)
\end{align*}
\]
Recursion

\[ w(n) = \sum_{m=0}^{n-1} w(m) w(n-1-m) \]

Solution

convolution \(\Rightarrow\) z-transform \(\Rightarrow\) quadratic equation

\[ \hat{w}(z) = \frac{z}{2} - \sqrt{\frac{z^2}{4} - zw(0)} \]

(\(w(0)\) still to be determined)
**Boundary condition**

\[
    w(0) = (\tilde{q} - q) \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} w(n_1) w(n_2) G(2(n_1 + n_2); q)
\]

**z space**

\[
    w(0) = \frac{\tilde{q} - q}{2\pi i} \oint \left\{ \frac{1}{2q} - \frac{1}{4q\sqrt{z}} \left[ \sqrt{(\sqrt{z} - 1)^2 - 4q} + \sqrt{(\sqrt{z} + 1)^2 - 4q} \right] \right\} \times \left( \frac{1}{2z} - \sqrt{\frac{1}{4z^2} - \frac{w(0)}{z}} \right)^2 dz.
\]

**Self-consistency equation for** \( w(0) \)
Evaluation the self-consistency equation

Molten phase
- \( w(0) = 0 \) is always a solution
- Interpretation: molten phase
- However: only know that \( Q(k, n)/z_1^N \rightarrow 0 \); do not know \( \Theta \)

Native phase
- Non-zero solution only if branch cuts do not overlap
- As \( w(0) \rightarrow 1/[4(1 + 2\sqrt{\tilde{q}})^2] \), integral converges to \( \approx 4(1 + 2\sqrt{\tilde{q}})^2 \)
- Non-zero solution only for \( \tilde{q} > \tilde{q}_c \approx q + 4(1 + 2\sqrt{\tilde{q}})^2 \)
- Interpretation: native phase

There is a phase transition at finite \( \tilde{q} \)
Open problems

- **Analytical argument** that high temperature phase is molten
- **Position** of phase transition (at $q = 4$ numerics says $\tilde{q}_c \approx 80$, analytical calculation says $\tilde{q}_c \approx 96$)
- **Order** of phase transition (numerically appears very weak)
Outline of part II

5 Boltzmann partition function

6 Molten RNA

7 Molten-native transition — hairpin

8 Molten-native transition — Cayley tree

9 Summary
Summary of part II

- The partition function of RNA can be calculated in polynomial time.
- Asymptotic behavior for homogeneous RNA can be calculated by analytical methods.
- The partition function of molten RNA has a characteristic $N^{-3/2}$ behavior.
- It takes a finite amount of sequence bias to enforce a native structure.
- Branchiness of the native structure strongly affects the molten-native phase transition.
Part III

Sequence disorder
Outline of part III

10 Introduction

11 Numerical evidence

12 Analytical arguments for glass transition

13 Properties of the glass transition

14 Summary
Outline of part III

10 Introduction

11 Numerical evidence

12 Analytical arguments for glass transition

13 Properties of the glass transition

14 Summary
Importance of sequence disorder

- Homogeneous models only good at high temperatures
- At low temperatures differences between bases matter
- Still want to know thermodynamic behavior
- Have to average over sequences
- Need quenched average

\[ F = -RT \log Z \]
Outline of part III

10 Introduction

11 Numerical evidence
   - Numerical method
   - Low temperature properties
   - Evidence for glass transition

12 Analytical arguments for glass transition

13 Properties of the glass transition

14 Summary
Zero temperature numerics

Observation
At \( T = 0 \) the free energy \( F \) equals the ground state energy \( E \).

Definition
Let \( E_{i,j} \) be the energy of the lowest energy structure for the RNA molecule starting at base \( i \) and ending at base \( j \).

Recursion equation
\[
E_{i,j} = \min_{k=i}^{j-1} \{ E_{i,j-1}, E_{i,k-1} + e(k,j) + E_{k+1,j-1} \}
\]

- Calculate from shortest to longest substrands
- \( O(N^3) \) algorithm for arbitrary sequence
- Repeat for many randomly chosen sequences
Zero temperature properties

Properties

- **Return probability** $p_{i,j} \equiv \Pr\{i \text{ and } j \text{ paired}\}$
  $\sim |i - j|^{-3/2}$ in molten phase

- **Size of structure** $\langle h \rangle = \sum_{i=1}^{N/2} \sum_{j=N/2+1}^{N} p_{i,j}$
  $\sim N^{1/2}$ in molten phase

Findings

- **Return probability**
  [Graph showing the return probability $p_{10}$ with a slope of $-1.34$.]

- **Size**
  [Graph showing the size $\langle h \rangle$ with a slope of $N^{0.7}$.]
  Ralf Bundschuh (Ohio State University) Statistical Physics of RNA folding
Is there a glass transition?

Observation
Thermodynamics at $T = 0$ different from molten phase.

Question
Is random RNA . . .

- . . . always molten for $T > 0$?
- . . . always glassy for $T < \infty$?
- . . . glassy below a finite glass transition temperature $T_c$ and molten above?
Finite temperature numerics

- Partition function numerics as a function of temperature
- Singularity in second derivative of specific heat

![Graph showing specific heat as a function of temperature](image)


- Very smooth transition at finite $T_c$
Outline of part III

10 Introduction

11 Numerical evidence

12 Analytical arguments for glass transition
   • Limit on the molten phase
   • Limit on the glass phase

13 Properties of the glass transition

14 Summary
Strategy

Objective
Show that there is a \textit{finite temperature} glass transition.

First step
Show that molten phase \textit{cannot be stable} for all $T > 0$.

Approach
Assume that molten phase is stable for all $T > 0$ and show that this is internally inconsistent.
Pinching energy

Object of interest

Pinching energy $\Delta F$ as a function of sequence length $N$.

Direct calculation

Molten phase stable for all $T$

$$\Rightarrow \quad G(T, N) \approx A(T)N^{-3/2} \exp[-f_0(T)N]$$

$$\Rightarrow \quad \Delta F = -RT \log \frac{G(T, N/2)G(T, N/2)}{G(T, N)} \approx \frac{3}{2}RT \log(N)$$
Lower bound

Best match

- Find match of exactly complementary pieces
**Lower bound**

**Best match**

- Find *match* of *exactly complementary* pieces
- Can be found as long as $N^2 4^{-\ell} > 1$
Lower bound

Best match

- Find match of exactly complementary pieces
- Can be found as long as $N^2 4^{-\ell} > 1$
- Longest one has length $\ell = \log N / \log 2$. 
Lower bound

Best match

- Find match of exactly complementary pieces
- Can be found as long as $N^{24^{-\ell}} > 1$
- Longest one has length $\ell = \log N / \log 2$

Pinch free energy

- Each base pair in exact match contributes $\varepsilon_0$
- Unpinched partition function
  $Z_{\text{unp}} \geq \exp\left[\frac{\varepsilon_0}{RT}\ell\right] \exp\left[-(N - 2\ell)f_0(T)\right]$
- Pinch free energy $\Delta F \geq [\varepsilon_0 + 2RTf_0(T)]\ell$
Putting it together

Combining everything

\[ \frac{3}{2} RT \geq [\varepsilon_0 + 2RTf_0(T)]/\log 2 \] (*)

Observation

\[ \varepsilon_0 + 2R \lim_{T \to 0} [Tf_0(T)] > 0 \]

Conclusion

(*) must break down at some temperature \( T_* \)
\[ \Rightarrow \] The molten phase cannot be stable below \( T_* \)
\[ \Rightarrow \] The glass transition must happen at a temperature above \( T_* \)
Small disorder

Second step
Show that molten phase persists down to some finite critical temperature.

Approach
Understand molten (no disorder) phase
⇒ introduce a little bit of disorder
⇒ Treat disorder as a perturbation

Model
Simplified model: $e(i,j)$ independent Gaussian random variables with mean $\varepsilon_0$ and variance $D$. 
Replica approach

\( n \) non-interacting RNAs in the same disorder (same “sequence”)

- Partition function of one molecule: \( Z \)
- Partition function of \( n \) molecules: \( Z^n \)
- Disorder average: \( \overline{Z^n} \)

Why care?

- Expansion of quenched free energy \(-RT\log Z\) in powers of disorder \( D \) starts as

\[
-RT\log Z = -RT \log \overline{Z} + \frac{RT}{2} \left[ \overline{Z^2} - \overline{Z}^2 \right] + O(D^2)
\]

\( \Rightarrow \) Small \( n \) replica averaged partition functions determine small disorder behavior
Replicated partition functions

How do the replicated partition functions look like?

- $\overline{Z^1}$ is molten phase partition function (homogeneous model)
- In general:
  $\overline{Z^n}$ is partition function of $n$ homogeneous RNA molecules that gain an extra energy $D$ whenever two molecules form a common bond.
Calculating replicated partition functions

Single replica
Molten phase ⇒ already calculated

Arbitrary $n$
Open problem, only approximate calculations exist, see next subsection

Two replicas
- Can be calculated exactly along the lines of what we did in part II
- Too long to present here ⇒ only rough idea
  (see me if you want to get the details)
Two replica calculation

- Order configurations of 2 replica system by common bonds

- Common bonds (---) form RNA structure themselves

- Represents sum over all possible choices of non-common bonds in the two replicas

- 1 replica → \( \ell^{-3/2} \) ⇒ 2 replicas → \((\ell^{-3/2})^2 = \ell^{-6/2}\)

- Effective picture: single RNA with “6-dimensional” loop entropies
Results from two replica calculation

Two replica results

- phase transition at finite $D_c$
- $D < D_c$:
  - Bubbles dominate
  - $\overline{Z^2} \sim N^{-3}$
  - 2 replicas fluctuate independently (molten)
- $D > D_c$:
  - Common bonds dominate
  - $\overline{Z^2} \sim N^{-3/2}$
  - 2 replicas have same configuration (glass)

Consequences for glass transition

Molten phase perturbatively stable
Outline of part III

10 Introduction

11 Numerical evidence

12 Analytical arguments for glass transition

13 Properties of the glass transition
   - Critical exponents
   - Physical picture

14 Summary
Properties of the glass transition

Field theory of the glass transition


Approach

- Replica partition function
- Expansion in strength of the disorder
- Renormalization through Operator Product Expansion
- Two loop calculation
- \( n \to 0 \) limit

Results

- Critical exponents the same at transition and in glass phase
- \( \langle h \rangle \sim N^{0.64} \), \( p_{i,j} \sim |i - j|^{-1.36} \) (very good agreement with numerics)
Picture of the glass transition


**Above glass transition:**
- Structures are largely *different*
- **Bubbles** of ground state structure appear
- Bubble size *diverges* as transition is approached

**Below glass transition**
- Structures are largely *locked* into ground state structure
- **Bubbles** of unlocked bases appear
- Bubble size *diverges* as transition is approached
Outline of part III

10 Introduction

11 Numerical evidence

12 Analytical arguments for glass transition

13 Properties of the glass transition

14 Summary
Summary of part III

- Random RNA is in a distinct glass phase at low temperatures
- A finite temperature glass transition separates the glassy from the molten phase
- The critical properties of the glass transition are perturbatively known through an RG calculation
- Exact results on the properties of the glass phase are still outstanding
Part IV

Force-extension experiments
Outline of part IV

15 Force-extension experiments

16 Quantitative modeling

17 Results for simple force-extension experiments

18 Nanopores

19 Summary
Outline of part IV

15 Force-extension experiments
   - Motivation
   - Experimental setup

16 Quantitative modeling

17 Results for simple force-extension experiments

18 Nanopores

19 Summary
Experimental methods to determine RNA structure

Experimental methods

- **X-ray crystallography**
- **Nuclear Magnetic Resonance**
- **Biochemical evidence**: protection essays
- **Correlated mutation analysis**

Problem

RNA is very floppy $\Rightarrow$ two RNA molecules rarely have the same three-dimensional structure

Idea

Look at **one** RNA molecule at a time.
Experimental setup

- Attach ends of RNA molecule to two beads
- Keep beads at fixed distance $R$ with optical tweezers
- Measure force $f$ on beads as function of distance $R$
Measurement

Subject

P5ab hairpin of *Tetrahymena thermophila* group I intron

Data

(Liphardt, Onoa, Smith, Tinoco, and Bustamante, *Science*, 2001)
Outline of part IV

15 Force-extension experiments

16 Quantitative modeling
   - Force from free energy
   - Secondary structures
   - Backbone
   - Putting it back together

17 Results for simple force-extension experiments

18 Nanopores

19 Summary
Calculating the force

Basic physics
energy = force \cdot distance

Force from free energy

Need free energy $F(r)$ at fixed end to end distance $r \Rightarrow$ force $f = \frac{\partial F(r)}{\partial r}$
Free energy from partition function

\[ F(r) = -RT \ln Z(r) \Rightarrow \text{need partition function } Z(r) \text{ at fixed distance } r \]

Partition function

\[ Z(r) = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } P \text{ with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[P]}{RT}} \]
Partition function II

\[ Z(r) = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } \mathcal{P} \text{ with distance } r \text{ and structure } S} e^{-\frac{E[S] + E[\mathcal{P}]}{RT}} \]
Partition function II

\[ Z(r) = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } \mathcal{P} \text{ with distance } r \text{ and structure } S} e^{-\frac{E[S] + E[\mathcal{P}]}{RT}} \]

\[ = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } \mathcal{P} \text{ of the “exterior bases” with distance } r \text{ and structure } S} e^{-\frac{E[S] + E[\mathcal{P}]}{RT}} \]
Partition function II

\[ Z(r) = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } P \text{ with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[P]}{RT}} \]

\[ = \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } P \text{ of the "exterior bases" with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[P]}{RT}} \]

\[ = \sum_{m=0}^{N} \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} \sum_{\text{polymer configurations } P \text{ of the "exterior bases" with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[P]}{RT}} \]
Partition function III

\[
Z(r) = \sum_{m=0}^{N} \sum_{\text{secondary structures } S} \sum_{\text{polymer configurations } \mathcal{P}} e^{-\frac{E[S] + E[\mathcal{P}]}{RT}}
\]

- \( Z(r) \) is the partition function.
- The sum over \( m \) runs from 0 to \( N \), where \( N \) is the number of secondary structures.
- Secondary structures are the polymer configurations of the "exterior bases" with distance \( r \) and structure \( S \).
- \( e^{-\frac{E[S] + E[\mathcal{P}]}{RT}} \) is the Boltzmann factor, representing the probability of the system being in a particular state.
Partition function $III$

\[
Z(r) = \sum_{m=0}^{N} \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} \sum_{\text{polymer configurations } \mathcal{P}} \sum_{\text{of the "exterior bases" with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[\mathcal{P}]}{RT}}
\]

\[
= \sum_{m=0}^{N} \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} e^{-\frac{E[S]}{RT}} \sum_{\text{polymer configurations } \mathcal{P}} \sum_{\text{of } m \text{ "exterior bases" with distance } r} e^{-\frac{E[\mathcal{P}]}{RT}}
\]
Partition function III

\[
Z(r) = \sum_{m=0}^{N} \left( \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} \sum_{\text{polymer configurations } \mathcal{P} \text{ of the "exterior bases" with distance } r \text{ and structure } S} e^{-\frac{E[S]+E[\mathcal{P}]}{RT}} \right)
\]

\[
= \sum_{m=0}^{N} \left( \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} e^{-\frac{E[S]}{RT}} \right) \left( \sum_{\text{polymer configurations } \mathcal{P} \text{ of } m \text{ "exterior bases" with distance } r} e^{-\frac{E[\mathcal{P}]}{RT}} \right)
\]

\[
= \sum_{m=0}^{N} \left( \sum_{\text{secondary structures } S \text{ with } m \text{ exterior bases}} e^{-\frac{E[S]}{RT}} \right) \left( \sum_{\text{polymer configurations } \mathcal{P} \text{ of ssRNA with } m \text{ bases with distance } r} e^{-\frac{E[\mathcal{P}]}{RT}} \right)
\]
Partition function IV

\[ Z(r) = \sum_{m=0}^{N} Q(m) W(m, r) \]

with

\[ Q(m) \equiv \sum_{\text{secondary structures } S} e^{-\frac{E[S]}{RT}} \]

and

\[ W(m, r) \equiv \sum_{\text{polymer configurations } \mathcal{P} \text{ of ssRNA}} e^{-\frac{E[\mathcal{P}]}{RT}} \]
Recursion equation

Reminder

\[ S_{i,j} = S_{i,j-1} + \sum_{k=i}^{j-1} Z_{i,k-1} e^{-\frac{e(k,j)}{RT}} Z_{k+1,j-1} \]

Definition

Let \( Q_j(m) \) be the partition function for the first \( j \) bases with exactly \( m \) exterior bases.

Generalization

\[ Q_j(m) = Q_{j-1}(m-1) + \sum_{k=1}^{j-1} Q_{k-1}(m) e^{-\frac{e(k,j)}{RT}} Z_{k+1,j-1} \]
Properties of recursion equation

\[ Q_j(m) = Q_{j-1}(m - 1) + \sum_{k=1}^{j-1} Q_{k-1}(m) e^{-\frac{e(k,j)}{RT}} Z_{k+1,j-1} \]

Properties:

- \( O(N^3) \) complexity
- \( Q(m) = Q_N(m) \)
- Can be easily generalized to Turner parameters
- Can be calculated by modifying the Vienna package (Hofacker, Fontana, Stadler, Bonhoeffer, Tacker, and Schuster, Monatshefte f. Chemie, 1994)
Polymer physics

Need

\[ W(m, r) = \sum \text{polymer configurations } \mathcal{P} \text{ of ssRNA with } m \text{ bases with distance } r \]

Model

- Elastic freely jointed chain
- Persistence length 1.9\(nm\)/base distance 0.7\(nm\)

Partition function

Calculation of \(W(m, r)\) is standard polymer physics.
Putting it together

\[ Q(m), \ W(m, r) \]

\[ Z(r) = \sum_{m=0}^{N} Q(m)W(m, r) \]

\[ F(r) = -RT \ln Z(r) \]

\[ f(r) = \frac{\partial F(r)}{r} \]

Web server

http://bioserv.mps.ohio-state.edu/rna
Outline of part IV

15 Force-extension experiments

16 Quantitative modeling

17 Results for simple force-extension experiments
   - Hairpin
   - A “real” molecule
   - Why the force-extension curve is smooth?

18 Nanopores

19 Summary
Apply to P5ab hairpin of *Tetrahymena thermophila* group I intron

Quantitative agreement!
The full group I intron

The group I intron of *Tetrahymena thermophila*

- Group I intron contains **pseudo-knot!**
- Quantitative modeling ignores pseudo-knot
  ⇒ known **inactive** conformation

Computational result

No sign of secondary structure!
What’s happening?

Intermediate structure

- Look at intermediate structure (here $r = 100 \text{nm}$)
- Like “Socks on the clothes line”
Results for simple force-extension experiments

Why the force-extension curve is smooth?

What’s happening?

Compensation effect

- Extension $r$ is increased
- One of the “socks” disappears
- The other “socks” take up the slag

⇒ No rapid change in force as sock disappears
⇒ smooth force-extension curve
Outline of part IV

15 Force-extension experiments

16 Quantitative modeling

17 Results for simple force-extension experiments

18 Nanopores
  • Introduction
  • Force-extension experiments

19 Summary
Nanopores

What is a nanopore?

A nanopore

- is a little **hole**
- is so small that **only single-stranded** but no double-stranded RNA can pass through it
- can be placed between two chambers such that there is only one nanopore connecting the chambers
Types of nanopores

natural ion channel (α-hemolysin) (Meller, J. Phys. Cond. Mat., 2003)

solid state pore (Storm et al., Nature Mat., 2003)
Combining nanopores with force-extension experiments

Suggestion

Combine a nanopore with a force-extension setup

Prediction

- The force will rise at every structural element
- The structural elements will break in their order along the sequence
Simulation approach

Simulation

- Fix $r(t)$ to be linear (set by experiment)
- **Degree of freedom**: number $n$ of base in the pore
- At each time $n$ can
  - increase:
    - $\rightarrow$ **calculable gain** in mechanical energy on the right
    - $\rightarrow$ potentially **loss** in binding energy calculable from $E[S]$
  - decrease:
    - $\rightarrow$ **calculable loss** in mechanical energy on the right

- Monte Carlo simulation (see also part V)
Apply to group I intron of *Tetrahymena thermophila*

- Signature of every structural element
- Can extract sequence position of stalling sites
Repeat pulling in opposite direction

Match stalling sites by sequence comparison ⇒ reconstruct structure
Overall structure reconstructed correctly
Can distinguish different structures on the same sequence
Even pseudoknot reconstructed
“Just” needs to be implemented experimentally
Outline of part IV

15 Force-extension experiments

16 Quantitative modeling

17 Results for simple force-extension experiments

18 Nanopores

19 Summary
Summary of part IV

- Quantitative description of single-molecule experiments possible
- Force-extension curves do not reveal secondary structure information due to compensation effects
- Single-molecule experiments reveal structure information with the help of a nanopore