Quantitative modeling of RNA single-molecule experiments

Ralf Bundschuh

Department of Physics, Ohio State University

Collaborators:
Ulrich Gerland, Universität zu Köln
Terence Hwa, UC San Diego

Outline:

• Introduction to RNA
• Single-molecule experiments
• Quantitative description of force-induced denaturation
• Why there is no structure in force-extension curves
• RNA structure kinetics
• Conclusions and outlook

Funding by NSF, DAAD, and ACS/PRF
Introduction to RNA

- RNA is heteropolymer of four different bases G, C, A, and U
- Primary structure: Sequence, e.g.,
  GCGGAUUAGCUCAUGGAGAGCGCCAGACUGAAAACUCUGGAGGUCCGUGUUCGAUCCACAGAAUUCGCACCA
- Strongest interaction: Watson-Crick base pairing (G–C and A–U)
  → secondary structure
- Spatial arrangement → tertiary structure
- Structure encoded in sequence
- Structure determines biological function
Single-molecule experiments

- RNA secondary structure difficult to determine experimentally since RNA is very floppy ⇒ no two copies of an RNA have the same shape
- Alternative: single-molecule experiments
  - 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 $R$
  - P5ab hairpin of tetrahymena thermophila group I intron

Liphardt, Onoa, Smith, Tinoco, and Bustamante, Science, 2001

- Method to learn something about secondary structure?
Quantitative prediction I

- Experiments hard $\Rightarrow$ quantitative modeling
- Two ingredients: secondary structure and polymer physics of backbone
- Backbone physics
  - Elastic freely jointed chain
  - Persistence length $1.9\text{nm}/\text{base distance }0.7\text{nm}$
  - Partition function at distance $R$ and length $m$: $W(R;m)$
- Secondary structure:
  - Need base pairing partition function $Q(m)$ of RNA molecule given $m$ “exterior” bases
- Total partition function:

$$Z(R) = \sum_{m=0}^{N} Q(m) W(R;m)$$

$\Rightarrow$ Complete knowledge of the thermodynamics
Quantitative prediction II

- How to get secondary structure partition function $Q(m)$ of RNA molecule given $m$ “exterior” bases?
- Consider all possible sets of base pairs formed
- Neglect pseudo-knots
- Diagramatic representation

- Assign energy $E[S] = \sum_{(i,j) \in S} \varepsilon_{i,j}$ to each structure $S$
- Total partition function

$$Z = \sum_m Q(m) = \sum_{\{S\}} \exp(-E[S]/T)$$

- Partition function generated exactly by Hartree equation

$$i \quad j \quad = \quad i \quad j-1 \quad j \quad + \sum_k i \quad k \quad k+1 \quad j-1 \quad j$$

- $O(N^3)$ algorithm for exact partition function (McCaskill, Biopolymers 29, 1990)
• As byproduct calculates partition function $Z_{i,j}$ for substrand from base $i$ to base $j$

• Introduce partition function $Q_j(m)$ for first $j$ bases with $m$ exterior bases

• Fulfills recursion equation

$$Q_j(m) = Q_{j-1}(m - 1) + \sum_{k=1}^{j-1} Q_{k-1}(m) e^{-\beta \varepsilon_{k,j}} Z_{k+1,j-1}$$

• Can be calculated by modifying “Vienna package”


• Uses very detailed free energy rules

Quantitative prediction IV

- Apply to P5ab hairpin of tetrahymena thermophila group I intron

- Quantitative prediction

- Disadvantages compared to experiment:
  - No tertiary structure or pseudo-knots
  - Parameters not exactly known

- Advantages over experiment:
  - Can be rapidly applied to arbitrary sequence
  - Intermediate structures can be investigated

- Offered as interactive web server

  http://bioserv.mps.ohio-state.edu/rna
FEC structure I

- Apply to full tetrahymena thermophila group I intron
- Group I intron contains pseudo-knot!
- Quantitative modeling ignores pseudo-knot ⇒ known inactive conformation
- No sign of secondary structure!

- Consistent with experiments on single-stranded DNA
  Maier, Strick, Croquette, and Bensimon, Single Molecules, 2000
**FEC structure II**

- What’s happening?
- Look at intermediate structure (here $R = 100nm$)
- “Socks on the clothes line”
  - 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
Structure by pulling I

- What can be done?

- Alternative experimental setup: pulling through a nano-pore

- Kinetic model: diffusion in a time dependent one-dimensional energy landscape ⇒ Monte Carlo simulation

- Signature of every structural element
- Extract sequence position of stalling sites
- Repeat experiment in reverse 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
• Nanopore experiments that are already done:

• DNA or RNA is pulled through pore via electric fields

• Translocation time monitored through reduction in counter ion current

• Only single-stranded DNA and RNA can go through pore

⇒ Coupling between structure and translocation kinetics

• Nanopore experiments ideal to study kinetics because time scale is set by experiment
• Simulation of RNA kinetics:

• Assume RNA structure rearrangements and pore translocation is dominated by energy barriers (breaks down at high fields)

⇒ Monte Carlo simulation

• Structure rearrangement:
  – base pair formation
  – base pair opening
  – base repairing

• Usual Metropolis rules with simple stacking energy and loop entropy

• Downhill move sets time unit (corresponds to about 1µs)
RNA kinetics III

- Movement through pore:
  - forward with rate $k_{\text{pore}} e^{E/2}$
  - backward with rate $k_{\text{pore}} e^{-E/2}$
  - Only allowed if new base in pore is unbound

- Speed through
  - “No rejection” algorithm choosing among all allowed moves according to their rates
  - Local reestimation of transition energies after each move
  - Data structures for $O(1)$ access to allowed moves
RNA kinetics IV

- Translocation time distributions

- For hairpin exponential (consistent with single barrier activated process)

- But: $\Delta G = 23.5 \rightarrow \exp(\Delta G) \approx 10^{11}$ while $\langle \tau \rangle \approx 86000$

- For random sequence distribution can be fit by Nelson-Lubensky distribution (for unstructured molecules !) Lubensky and Nelson (1999)

- However, average velocity for random sequence larger than for corresponding homopolymer
Why is translocation diffusion-like and fast? → look at translocation

Major barrier to translocation is entry of structure into the pore

Once pore is embedded in structure translocation resembles diffusion in largely flat landscape

Reminiscent of equilibrium free energy landscape

How to calculate effective free energy landscape?

Partition function $Z_{i,j}$ for RNA strand from base $i$ to base $j$

Free energy with pore at base $i$ given by “pinching energy”

$$F_i = -\log \frac{Z_{1,i-1}Z_{i+1,N}}{Z_{1,N}}$$
Conclusions and outlook

Conclusions:

- **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 and kinetic information with the help of nano-pores**

Challenges:

- **Comparison with experiments on medium-sized RNA**

- **Include protein–RNA interactions**

- **Quantitative kinetics**