Bayesian analysis of DNA unfolding in nanochannels

Michael A. Lomholt
University of Southern Denmark, Odense, Denmark

Work with (by):

Jens Krog, Odense, Denmark
Mohammadreza Alizadehheidaria, Chalmers, Gothenburg, Sweden
Erik Werner, Gothenburg U., Sweden
Jonas Tegenfeldt, Lund, Sweden
Joachim Fritzche, Chalmers, Gothenburg, Sweden
Bernhard Mehlig, Gothenburg U., Sweden
Fredrik Westerlund, Chalmers, Gothenburg, Sweden
Tobias Ambjörnsson, Lund, Sweden

Warning: work in progress
Outline

• DNA unfolding experiments
• Bayesian statistics
• Stochastic model of the DNA unfolding
• Results and theoretical estimates
DNA unfolding experiment

Hairpin lengths vs. time

[Graph showing hairpin lengths vs. time]
The average behavior

Simple modeling: \[
\frac{d\langle x \rangle}{dt} = -\frac{f_U}{\gamma_0} \langle x \rangle
\]

Depends on: \[
\frac{f_U}{\gamma_0}
\]
Bayesian inference

Models: $M_1, M_2, \ldots$

Bayes formula: $P(M_i|\text{data}) = \frac{P(\text{data}|M_i)P(M_i)}{P(\text{data})}$

Model selection: $\frac{P(M_i|\text{data})}{P(M_j|\text{data})} = \frac{P(\text{data}|M_i)P(M_i)}{P(\text{data}|M_j)P(M_j)}$

Model evidence: $P(\text{data}|M_i) = \int P(\text{data}|\boldsymbol{\theta}, M_i)P(\boldsymbol{\theta}|M_i)d\boldsymbol{\theta}$

Posterior probability for the parameters: $P(\boldsymbol{\theta}|\text{data}, M_i) = \frac{P(\text{data}|\boldsymbol{\theta}, M_i)P(\boldsymbol{\theta}|M_i)}{P(\text{data}|M_i)}$
Nested Sampling

\[ Z = P(\text{data}|M_i) = \int P(\text{data}|\theta, M_i)P(\theta|M_i)d\theta \]

Calculated via Nested Sampling
[John Skilling, Bayesian Analysis 1, 833 (2006)]
Now, for a polymer in an unconfined medium the Rouse segment an unfolding force inside a channel as seen in Fig. 1 (bottom). On each appeared we are left with one hairpin. The dynamics proceeds is photo-cut at a random point along its contour. There is an order bead size), the hydrodynamic interactions between confined in a very narrow tube (channel diameter of the friction scales linearly with the length of the polymer.

We consider the later stage first.

Let us now describe the equations of motion. We de-

Assume: $\mu = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ with $\mu_1 = \frac{1}{\gamma_0 (X - x)}$ & $\mu_2 = \frac{1}{\gamma_0 x}$

and $x = x_3 - x_2 = \frac{1}{2} (X + x_1 - x_2)$

Total length: $X = (x_3 - x_2) + (x_3 - x_1)$ assumed fixed

Fokker-Planck equation for $p(x_1, x_2, t)$: $\partial_t p = -\nabla^T (\mu f p - k_B T \mu \nabla p)$
The likelihood function

Is calculated via the Langevin equation:

\[
dx = \left[ -\frac{2f_U}{4} \left( \frac{1}{\gamma_0(X - x)} + \frac{1}{\gamma_0x} \right) + \frac{k_BT}{4} \frac{d}{dx} \left( \frac{1}{\gamma_0(X - x)} + \frac{1}{\gamma_0x} \right) + \eta_x(t) \right] dt
\]

- Drift
- Term arising due to x-dependent friction
- Noise

Covariance of noise:

\[
\langle \eta_x(t)\eta_x(t') \rangle = \frac{k_BT}{2} \left( \frac{1}{\gamma_0(X - x)} + \frac{1}{\gamma_0x} \right) \delta(t - t')
\]

Gaussian
Steps with: mean \( \overline{\Delta x} = \left( -\frac{2f_U}{4} \mu_+ + \frac{k_BT}{4} \partial_x \mu_+ \right) \Delta t \) where \( \mu_+ = \frac{1}{\gamma_0(X - x)} + \frac{1}{\gamma_0x} \)

variance \( \sigma^2 = \left\langle (\Delta x - \overline{\Delta x})^2 \right\rangle = \frac{k_BT}{2} \mu_+ \Delta t \)
Results

\[ f_U = (0.027 \pm 0.004) \text{ pN} \]
\[ \gamma_0 = (2.4 \pm 0.3) \text{ mPa s} \]

\[ f_U = (0.019 \pm 0.003) \text{ pN} \]
\[ \gamma_0 = (1.8 \pm 0.2) \text{ mPa s} \]

Estimate of force

\[ f_U \approx \frac{k_B T}{\text{length scale}} \approx \frac{4 \text{ pN nm}}{100 \text{ nm}} = 0.04 \text{ pN} \]
Estimate of friction

\[ f = \gamma_0 v \]

\[ \gamma_0 = \frac{2\pi}{\ln\left(\frac{r_2}{r_1}\right)} \eta \]

Estimate:

\[ \frac{r_2}{r_1} \approx 50 \& \eta \approx 1 \text{ mPa s} \implies \gamma_0 \approx 1.6 \text{ mPa s} \]
Model selection

friction = $\gamma_0 x^\nu$

log $Z_{\nu=1} = -173.1$

log $Z_{\nu \in [0,2]} = -177.6$

log $Z_{\nu=0} = -185.8$

log $Z_{\nu \in [0,2]} = -248.8$

log $Z_{\nu=0} = -253.9$

log $Z_{\nu=1} = -255.3$
Conclusions

Bayesian analysis

• is conceptually simple

• puts focus on precise modeling of data

• but it can be computationally demanding

• (can use nested sampling)
Collaborators

**Jens Krog, Odense, Denmark (Bayesian analysis)**
Mohammadreza Alizadehheidaria, Chalmers, Gothenburg, Sweden
Erik Werner, Gothenburg U., Sweden
Jonas Tegenfeldt, Lund, Sweden
Joachim Fritzsche, Chalmers, Gothenburg, Sweden
Bernhard Mehlig, Gothenburg U., Sweden
Fredrik Westerlund, Chalmers, Gothenburg, Sweden
Tobias Ambjörnsson, Lund, Sweden