Topic 5: Random polymers – DNA and polymer elasticity
Overview:

We’ll see that a long polymer can be viewed as a random walk.

Long DNA polymers like our chromosomes do random walks.

Random polymers have elasticity – entropic elasticity.

Applications: DNA breathing, stretching random polymers.
Imagine a chain made up of links of size, a

Each link can be oriented randomly

Links do not interact – i.e. there is no self-avoidance → all conformations have $E = 0$
Real biopolymers as freely jointed chains

To discretize a real polymer into links, we consider ‘a’ to be the length over which the polymer is effectively rigid.

We’ll see that this can be defined exactly in terms of a measurable quantity called persistence length.
Each segment has equal probability of going left or right.

Shown are possible ways of winding up with a certain number of steps to the right, $n_r$.

On average what is the end-to-end distance?

let’s calculate this …
End-to-end distance

What is mean displacement?

\[ \langle R \rangle = \langle \sum_{i=1}^{N} x_i \rangle = \sum_{i=1}^{N} \langle x_i \rangle = 0 \text{ since } \langle x_i \rangle = 0 \]

Variance?

\[ \langle R^2 \rangle = \langle \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \rangle = \sum_{i=1}^{N} \langle x_i^2 \rangle + \sum_{i \neq j}^{N} \langle x_i x_j \rangle \]

For the 2nd term \( \langle x_i x_j \rangle = 0 \) since the steps are independent and there are equal # of \(+a^2\) and \(-a^2\) terms.

1st term \( \langle x_i^2 \rangle = a^2 \) so

\[ \langle R^2 \rangle = Na^2 \]

or RMSD

\[ \sqrt{\langle R^2 \rangle} = a \sqrt{N} \]

So for a random walk, on average the spread of the polymer goes as \( \sqrt{N} \)
Probability of configurations:

- What is the probability of a given configuration?
  - Walk characterized by \( n_r \) steps to the right and \( n_l = N - n_r \) steps to the left.

- For a polymer of length \( N \), how many configurations have exactly \( n_r \) steps?

  \[
  \text{# of configs} = W(n_r, N) = \frac{N!}{n_r!(N-n_r)!}
  \]

  \( W = 3, n_r = 2 \)

  \( \equiv \text{# of unique combinations of } n_r \text{ steps out of } N \)

  \( \text{e.g. } N = 3 \ \& \ n_r = 1 \)

  \( W(1,3) = 3 \)
Probability of configurations:

For a given configuration, since \( r = \frac{1}{2} \) and there are \( N \) steps, the prob = \( \left( \frac{1}{2} \right)^N \)

So the total prob for configurations with \( n_r \) steps

\[
P(n_r; N) = \frac{N!}{n_r!(N-n_r)!} \left( \frac{1}{2} \right)^N
\]

This is the same as the binomial distribution := prob of seeing \( n_r \) outcomes in \( N \) events where \( p = \frac{1}{2} \)

shorthand:

\[
\text{binom}(n_r; N) = \binom{N}{n_r} \left( \frac{1}{2} \right)^n_r \left( \frac{1}{2} \right)^{N-n_r}
\]

and for \( p = \frac{1}{2} \)

\[
= \frac{N!}{n_r!(N-n_r)!} \left( \frac{1}{2} \right)^N
\]

mean of binomial:

\[
\mu = \langle n_r \rangle = N \cdot p_r = \frac{N}{2} \text{ for } p_r = \frac{1}{2}
\]

\[
\sigma^2 = N \cdot p_r (1 - p_r) = \frac{N}{4} \text{ for } p_r = \frac{1}{2}
\]
Proof that mean = N \cdot p

\[ \langle n_r \rangle = \sum_{n_r=1}^{N} n_r \frac{N!}{n_r! (N-n_r)!} p_r^{n_r} (1-p_r)^{N-n_r} \]

\[ \Rightarrow \sum_{n_r=1}^{N} \frac{N \cdot (N-1)!}{n_r! (n_r-1)! (N-n_r)!} p_r p_r^{n_r-1} (1-p_r)^{N-n_r} \]

\[ = N \cdot p_r \sum_{n_r=1}^{N} \frac{(N-1)!}{n_r-1! (N-n_r)!} p_r^{n_r-1} (1-p_r)^{N-n_r} \]

\[ = 1 \Rightarrow \sum p(n_r) = 1 \]

= N \cdot p_r
Binomial distribution becoming Gaussian

For small $N$, the binomial distribution is asymmetric.

For large $N$, binomial $\rightarrow$ Gaussian

So $P(n_r; N) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(n_r - \mu)^2}{2\sigma^2}\right)$

where $\mu = \frac{N}{2}$ and $\sigma^2 = \frac{N}{4}$
End-to-end probability distribution

Let’s turn this into a distribution of $R$ = end-to-end distance:

$$ R = n_{-}a - n_{+}a = n_{-}a - (N-n_{-})a $$

$$ = 2n_{-}a - Na $$

Since $n_{-}$ is a Gaussian so will $R$

Mean: $\langle R \rangle = 2a \langle n_{-} \rangle - Na = 2aN - Na = 0$

Variance = $\langle R^2 \rangle = \langle (2n_{-}a - Na) (2n_{-}a - Na) \rangle$

$$ = 4a^2 \langle n_{-}^2 \rangle - 4Na^2 \langle n_{-} \rangle + Na^2 $$

Now:

$$ \langle n_{-}^2 \rangle = \sigma^2 + \mu^2 = \frac{N}{4} + \frac{N^2}{4} $$

So:

$$ \langle R^2 \rangle = 4a^2 \left( \frac{N}{4} + \frac{N^2}{4} \right) - 4Na^2 \langle n_{-} \rangle + Na^2 $$

$$ = Na^2 $$ as before

So:

$$ \mathcal{P}(R) = \frac{1}{\sqrt{2\pi \sigma_R^2}} \exp \left( -\frac{R^2}{2\sigma_R^2} \right) $$

where $\sigma_R^2 = N \sigma^2$

Note: the most likely end-to-end distance, $R = 0$

If you pull on the polymer, it will want to restore $R \to 0$

purely entropic force
Persistence length

At what length, $\xi_p$, does the polymer become uncorrelated?

Correlation between tangent vectors obeys the following relation:

$$\langle T(s) \cdot T(u) \rangle = e^{-\frac{|s-u|}{\xi_p}}$$

For DNA, $\xi_p = 50 \text{ nm} \approx 150 \text{ bp}$
Persistence length and link size

- How does $\xi_p$ relate to random walk length $a$?

$$<R^2> = \left< \int_0^L ds \bar{E}(s) \cdot \int_0^L du \bar{E}(u) \right>$$

$$= 2 \int_0^L ds \int_0^L du \ e^{-(u-s)/\xi_p}$$

- Now take $L \gg \xi_p$ and let $x = u-s$, so

$$<R^2> \approx 2 \int_0^L ds \int_0^\infty e^{-x/\xi_p} = 2L \xi_p$$

- Compare with our previous result $<R^2> = N \xi = (Na)^2$.

So

$$a = 2 \xi_p$$

- Thus when modelling DNA as a freely jointed chain the segments should have a length

$$a = 2 \xi_p = 2(50 \text{ nm}) = 100 \text{ nm} \approx 300 \text{ bp}$$

$a = 300 \text{ bp}$ ⇒ Random walk for DNA
Sizes of genomes: Radius of gyration

log-log plot

slope = \frac{1}{2}

so chromosomes are like random poly when they are free in solution

Compare with E. coli size \sim 2 \mu m

A ok estimate is \langle R_g^2 \rangle = \sqrt{2L^2/3}

Better estimate is radius of gyration

\langle R_g^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} (R_i - \bar{R}_m)^2

Can show

\sqrt{\langle R_g^2 \rangle} = \frac{L^{2/3}}{3} \approx \frac{1}{3} \sqrt{N_{bp}^{2/3}} \text{ nm}

# of bp
Applications:

(A) DNA bubble

(B) RNA hairpin

(C) looping by transcription factor

(D) chromosome

Figure B.18 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
DNA melting and PCR:

When DNA is in solution it is continually opening and closing, and at high enough temperatures it will completely open, i.e. melt.

Q: At a given temperature, what is the typical bubble size in basepairs?
DNA melting and PCR: 1D random walk

- The free energy cost (entropic cost) of bubbles can be modeled using a random walk model for the single stranded DNA.
- Can use model to calculate melting temperature.
Consider the free energy difference between the closed and bubble states:

\[ F_c \quad \quad F_c + nE - TS_{\text{bubble}} \]

So,

\[ \Delta F = (F_c + nE - TS_{\text{bubble}}) - F_c = nE - T \frac{k_B}{2} \ln W_{\text{bubble}} \]

Here, \( E > 0 \) is the energy associated with pairing a base pair.

Note,

\[ W_{\text{bubble}} = \sum_0^\infty (n) \cdot (N-n) \]

\[ \text{# of places to put a bubble of size } n \]

\[ \text{# of copies of } 2 \text{ ssDNA of length } n \]

\( n \)

\[ \text{closed loop of length } 2n, \text{ returns to } x=0 \]
DNA melting and PCR: Entropy of loop

\[ n_r = n \Rightarrow n_l = n \quad \& \quad x = 0 \]

\[
S_0(n) = \frac{(2n)!}{n! (2n-n)!} = \frac{(2n)!}{n! n!}
\]

use \[ n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \equiv \text{Stirling} \quad \text{for } n \gg 1 \]

Then

\[
S_0(n) = \frac{2^{2n}}{\sqrt{\pi n}}
\]

so

\[
\frac{\Delta F}{k_B T} = n \varepsilon - \frac{1}{2} \ln \left( \frac{2^{2n}}{\sqrt{\pi n}} \right)
\]

\[
= n \left( \varepsilon - 2 \ln 2 \right) + \frac{1}{2} \ln n - \ln (N-n)
\]

\[
\approx \frac{1}{k_B T} E + \frac{1}{2} \ln \frac{n}{(N-n)}
\]
DNA melting and PCR: Free energy minimization

- At what value $n^*$ does $\Delta F$ take a minimum?
- Solve for $n^*$ from $\frac{d(\Delta F)}{dn} = 0$

$$\Rightarrow (\varepsilon - 2 \ln 2) + \frac{1}{2n} + \frac{1}{N-n} = 0$$

Solution:

$$n^* = N \left[ \frac{1 + \Delta \varepsilon - \sqrt{1 + 6 \Delta \varepsilon + \Delta \varepsilon^2}}{\Delta \varepsilon} \right]$$

where $\Delta \varepsilon = 2 (\varepsilon - 2 \ln 2)$

- Competition between energy & entropy of breaking one base pair
- For $\Delta \varepsilon > 0$, no solutions with $n^* < N \Rightarrow$ no bubbles

$$\Rightarrow \frac{E}{k_B T} = 2 \ln 2 > 0$$

or

$$k_B T < \frac{E}{2 \ln 2} \quad \text{for no bubbles}$$

- For $\Delta \varepsilon < 0$ or $k_B T > \frac{E}{2 \ln 2}$, bubbles are possible
DNA melting and PCR: equilibrium bubble size

PCR works by cycling temperature over this melting transition
If we apply a force to a random polymer to stretch it, what will its force vs extension characteristic look like?

Will it be like Hook’s law?

Using an optical trap (or AFM) we can pull on DNA, proteins, RNA to measure how they stretch.

These experiments will allow us to determine the persistence length of these polymers directly at the single molecule level.
Pulling on a multidomain protein

Our 1D random poly model

\[ L_{tot} = N \alpha \]
Pull on polymer with force, $f$, what will average extension $\langle L \rangle$ be?

Free energy of random chain:

$$F = -k_B T \ln W(L, L_{tot})$$

In trap, polymer has also done work, $-FL$

So

$$F_t = -FL - k_B T \ln W(L, L_{tot})$$

Instead of $L$, consider # of right steps, $n_r$, then

$$W(n_r; N) = \frac{N!}{n_r! (N-n_r)!}$$

Stirling's

So

$$F(n_r) = -fn_r \alpha + k_B T \left[ n_r \ln n_r + (N-n_r) \ln (N-n_r) \right]$$
Free energy of stretched freely jointed chain: equilibrium length

For a given applied force, what will the equilibrium stretch be?

Equilibrium \( n_r^* \Rightarrow \frac{dF}{dn_r} = 0 \)

\[ -f_a + k_B T \ln n_r - k_B T \ln n_L = 0 \]

or \( n_r^* = e^{f_a/k_B T} \)

\( \% \) extension \( z = \frac{\langle L \rangle}{\langle L_{tot} \rangle} = \frac{n_r - n_L}{n_r + n_L} = (\frac{n_r}{n_L} - 1)(\frac{n_r}{n_L} + 1) \)

so \( z = \tanh \left( \frac{f_a}{k_B T} \right) \)
Comparison to data:

- FJC does OK for small forces. Better model is worm-like chain which takes into account self avoidance and bending energy.

- Small force limit: \( \frac{f a}{k_0 T} \ll 1 \)

or \( 2 f a \ll k_0 T \) or \( f \ll \frac{k_0 T}{2a} \)

for dsDNA: \( a = 50 \text{nm} \Rightarrow f \ll 40 \text{fN} \)

for ssDNA: \( a = 0.75 \text{nm} \Rightarrow f \ll 3 \text{pN} \)
Hook's law limit at low forces:

- Taylor expand $z = \tan h(x)$; $x = 2f^2 / k_BT$
  \[
  \Rightarrow \quad <L> = \left[ \frac{L_0 + 2f}{k_BT} \right] f \quad \text{cf } x = \frac{f}{k}
  \]

- Effective spring: $k = \frac{k_BT}{2k + \frac{2}{9}} = \text{entropic spring}$
  
  for small forces

\[\text{ssDNA} \quad \downarrow \quad \text{dsDNA} \quad \text{softer spring} \quad \rightarrow f\]

for $\lambda$-phage: $L_{tot} = 16 \, \mu m$

- ssDNA \quad $k_s = 160 \, \text{fN/\mu m}$
- dsDNA \quad $k_{ds} = 2.3 \, \text{fN/\mu m}$

- Entropic forces always depend on temperature.
- For a random coil, higher $T \Rightarrow$ stiffer spring.
Summary:

- Long (unfolded) polymers can often be treated as a random walk
- Size of random polymer goes as $\sim \sqrt{N}$
- Most configurations have $R = 0$
- Leads to an entropic force that resists stretching
- DNA melting is a balance between entropy and binding energy
- The stretching of a random polymer is like a spring at low forces
- Can determine the persistence length of the polymer
- Stretching data is not well fit at larger forces