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

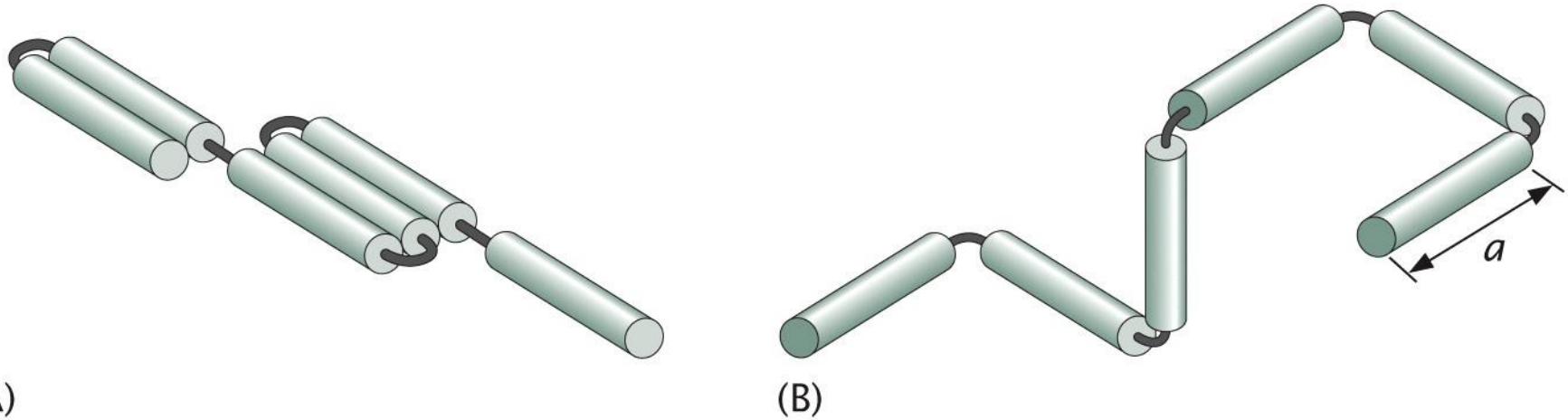Links do not interact – i.e. there is no self-avoidance
→ all conformations have E = 0

(A)

(B)

Figure 8.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

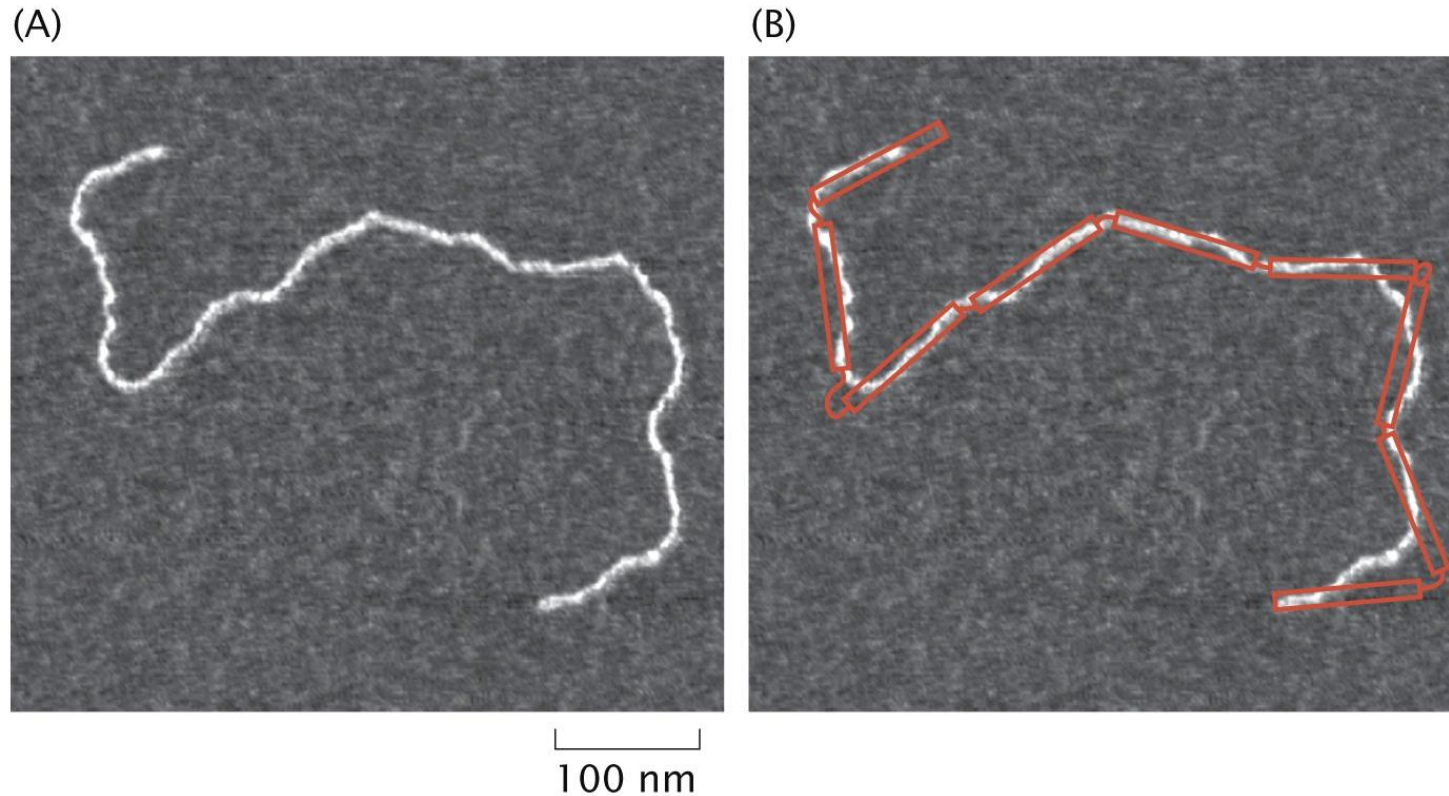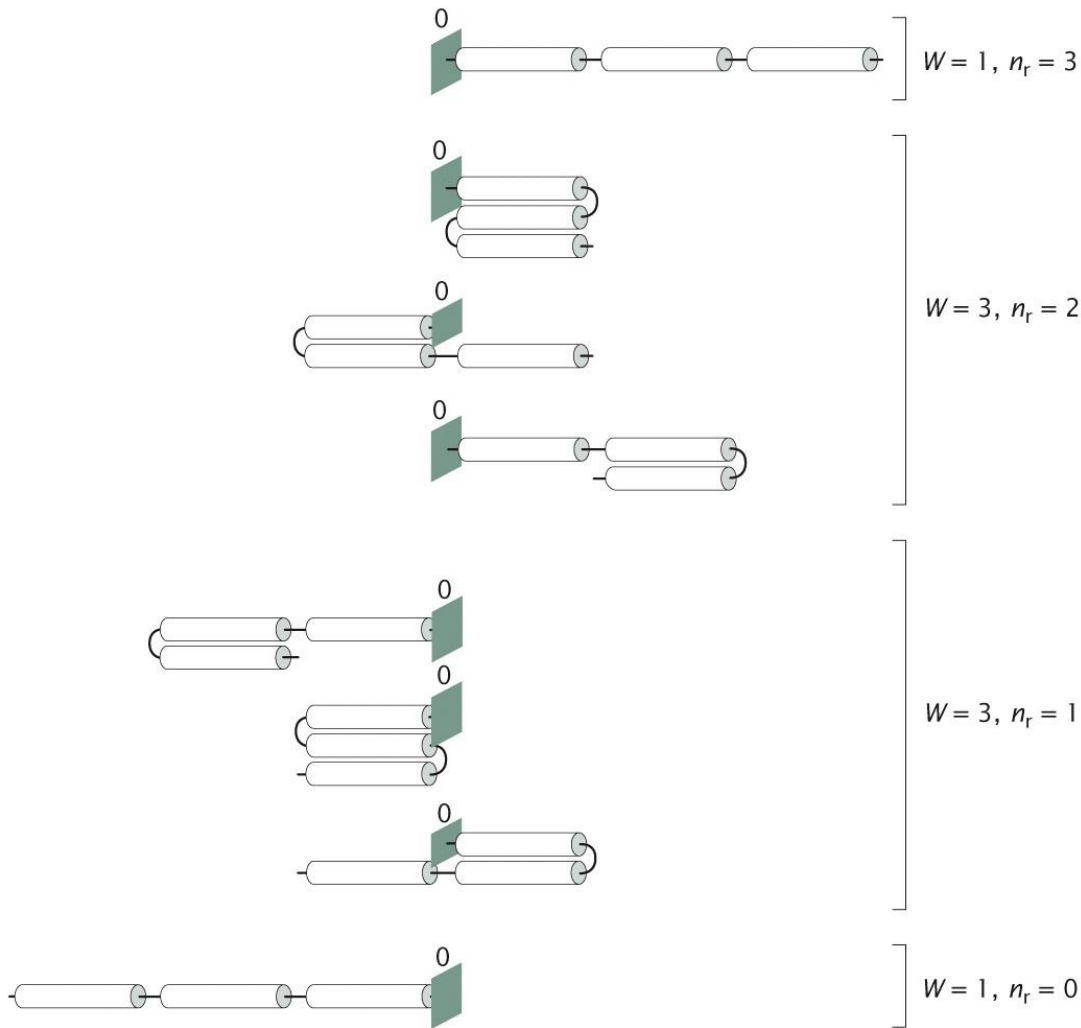# Real biopolymers as freely jointed chains

(A)   (B)

100 nm

Figure 8.2 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

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

$W = 1$, $n_r = 3$

$W = 3$, $n_r = 2$

$W = 3$, $n_r = 1$

$W = 1$, $n_r = 0$

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 …

Figure 8.3 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

- What is mean displacement?

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

- 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 here are equal # of $+a^2$ and $-a^2$ terms.

- 1st term $\langle x_i^2 \rangle = a^2$ so

$$\langle R^2 \rangle = N a^2$$

- or RMSD

$$\boxed{\sqrt{\langle R^2 \rangle} = a \sqrt{N}}$$

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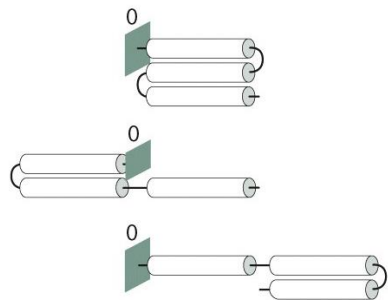
Probability of configurations:

- 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.

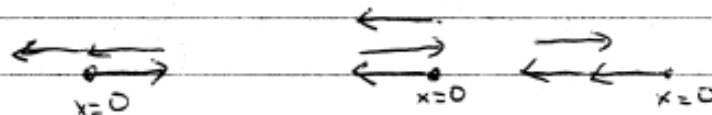- For a polymer of length N, how many configurations have exactly n_r steps?

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

$W = 3,\ n_r = 2$

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

- e.g. $N = 3$ & $n_r = 1$

$$W(1,3) = 3$$

$x = 0$ $x = 0$ $x = 0$

Probability of configurations:

- For a given configuration, since $p_r = p_L = \frac{1}{2}$ 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 $\equiv$ prob of seeing n_r outcomes in N events where $p_r = q$

$$\text{binom}(n_r; N) = \binom{N}{n_r} p_r^{n_r} (1-p_r)^{N-n_r}$$

shorthand

and for $p_r = \frac{1}{2}$

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

mean of binomial:

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

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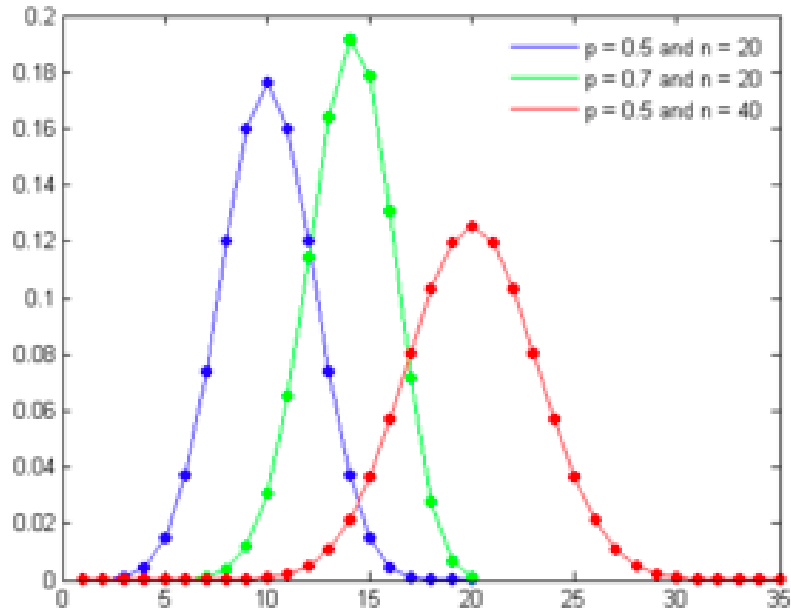
Proof that mean = N p

Proof:

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

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

$$= N \, p_r \sum_{n_r=1}^{N} \frac{(N-1)!}{(n_r-1)!(N-n_r)!} p_r^{n_r-1} (1-p_r)^{N-n_r}$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{=1 \;\Rightarrow\; \sum p(n_r) = 1}$$

$$= N \cdot p_r$$

Binomial distribution becoming Gaussian

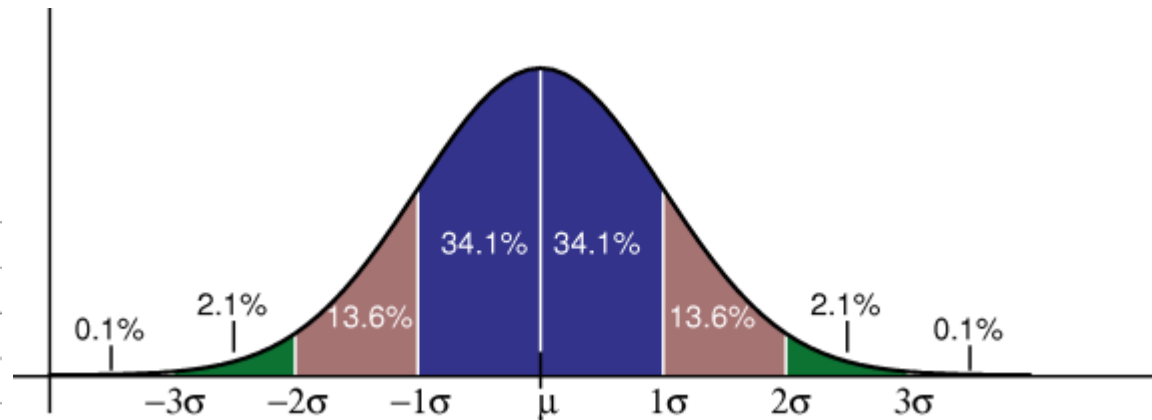

For small N, the binomial distribution is asymmetric

For large N, binomial → 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 $\mu = \dfrac{N}{2}$ and $\sigma^2 = \dfrac{N}{4}$

End-to-end probability distribution

• Let's turn this into a distribution of $R \equiv$ end to end distance.

$$R = n_r \cdot a - n_L a = n_r a - (N - n_r)a$$

$$= 2 n_r a - N a$$

• Since n_r is a gaussian so will R

mean $\equiv \langle R \rangle = 2a \langle n_r \rangle - Na = 2a \frac{N}{2} - Na = 0$

variance $\equiv \langle R^2 \rangle = \langle (2n_r a - Na)(2n_r a - Na) \rangle$

$$= 4a^2 \langle n_r^2 \rangle - 4Na^2 \langle n_r \rangle + Na^2$$

now

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

so

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

$$= Na^2 \quad \text{as before}$$

so

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

where $\sigma_R^2 = Na^2$

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

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

purely entropic force

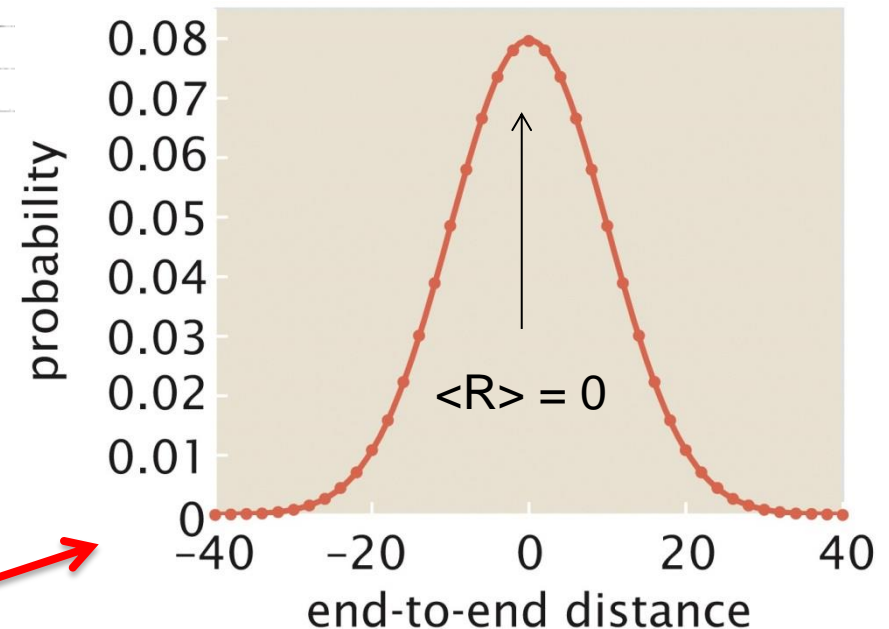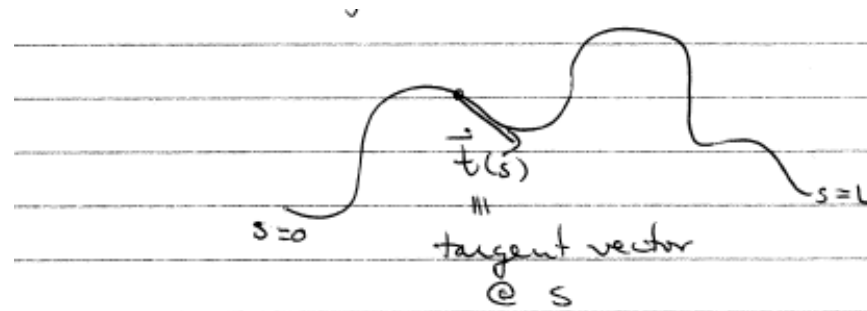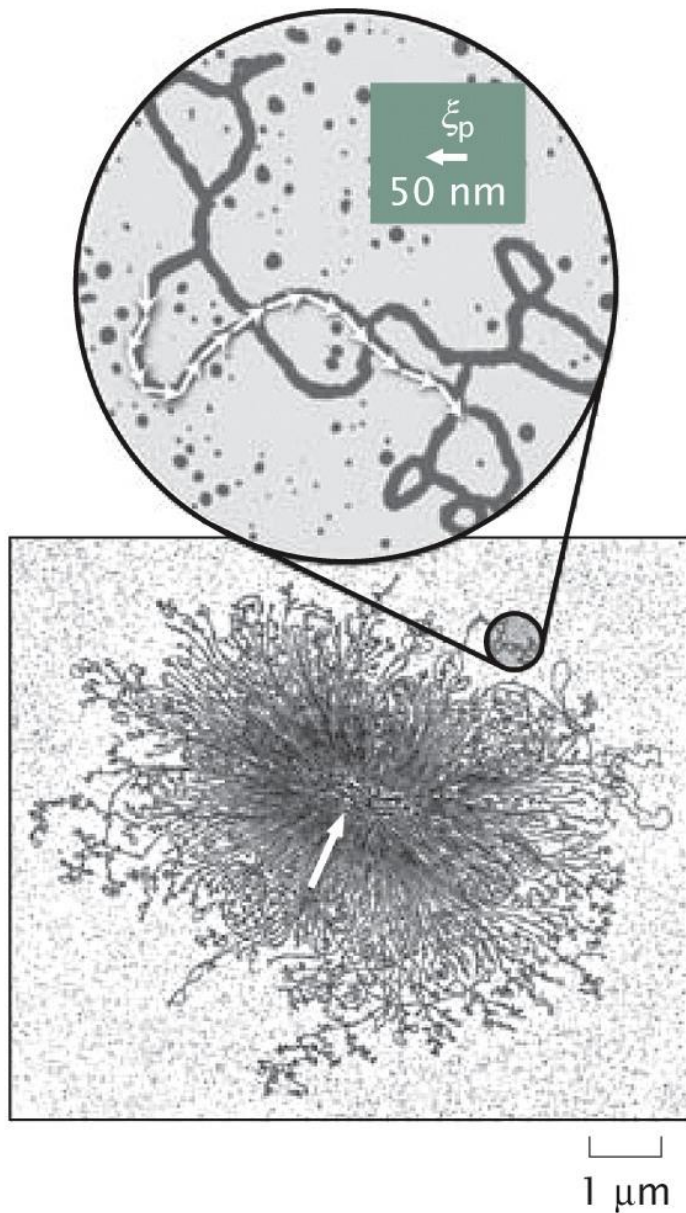$\langle R \rangle = 0$

Figure 8.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

50 nm

1 μm

Figure 8.5 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

$\vec{t}(s)$

$s=0$

$s=L$

tangent vector @ s

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

- Correlation between tangent vectors obeys the following relation

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

- For DNA, $\xi_p = 50\,nm \approx 150\,bp$

- How does ξ_p relate to random walk length $= a$

$$\langle R^2 \rangle = \left\langle \int_0^L ds\, \vec{t}(s) \cdot \int_0^L du\, \vec{t}(u) \right\rangle$$

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

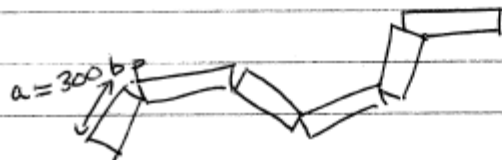- now take $L \gg \xi_p$ and let $x = u - s$, so

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

- Compare with our previous result, $\langle R^2 \rangle = N a^2 = (Na)a$
$$\underbrace{}_{L}$$

So
$$a = 2\xi_p$$

- Thus when modelling DNA as a freely jointed chain the segments should have a length
$a = 2\xi_p = 2(50\,nm) = 100\,nm \approx 300\,bp$

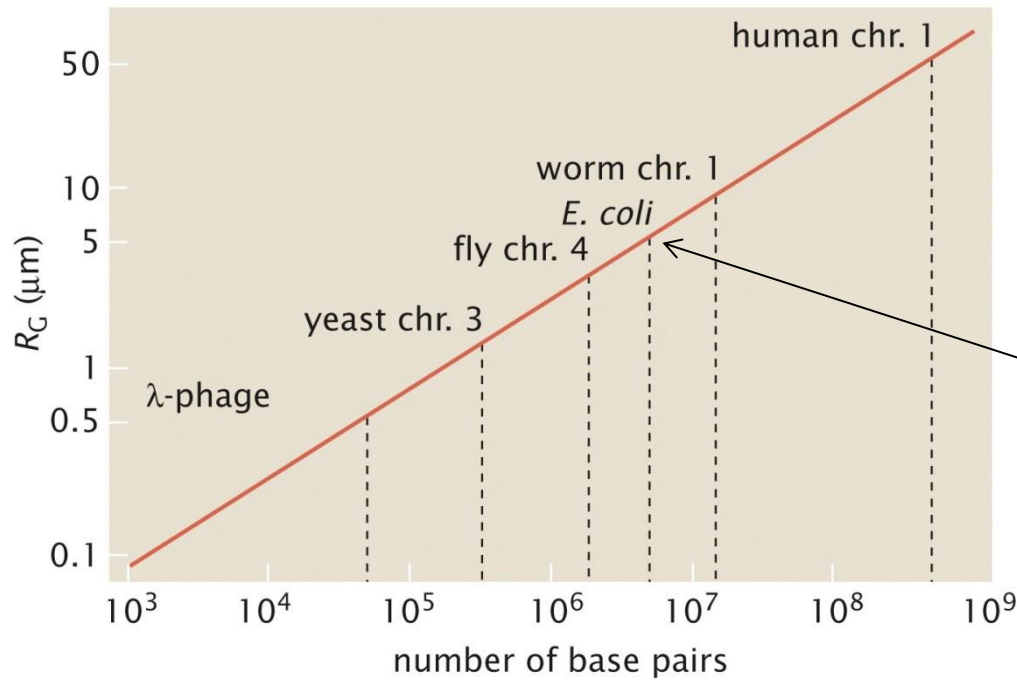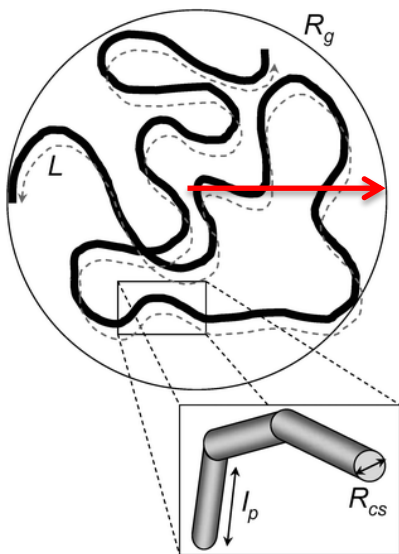$a = 300\,bp$ $\equiv$ Random walk for DNA

Sizes of genomes: Radius of gyration

log-log plot

slope = ½

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

Compare with E.coli
size ~ 2 μm

An ok estimate is $\sqrt{\langle R^2 \rangle} = \sqrt{2L\frac{\xi_p}{2}}$

• Better estimate is radius of gyration

$$\langle R_G^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle (\vec{R_i} - \vec{R_{cM}})^2 \rangle$$
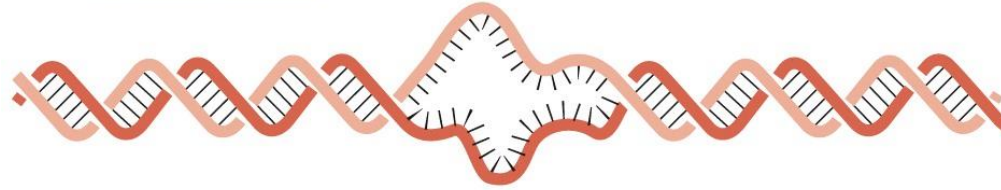
center of mass

can show

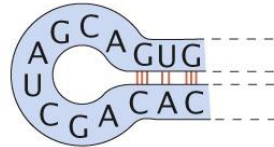$$\sqrt{\langle R_G^2 \rangle} = \sqrt{\frac{L\frac{\xi_p}{3}}{3}} \approx \frac{1}{3}\sqrt{N_{bp}\,\xi_p}\ \ nm$$
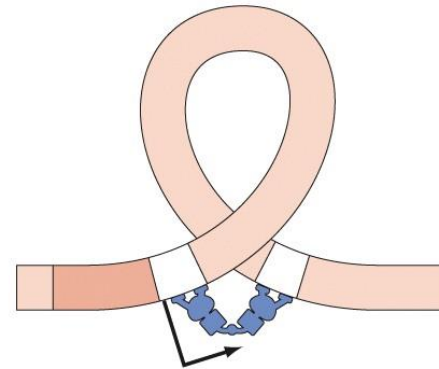
# of bp

Applications:

(A) DNA bubble

(B) RNA hairpin

(C) looping by transcription factor
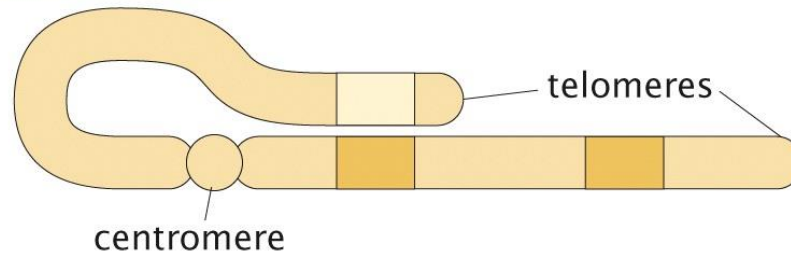
(D) chromosome

telomeres

centromere

Figure 8.18 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
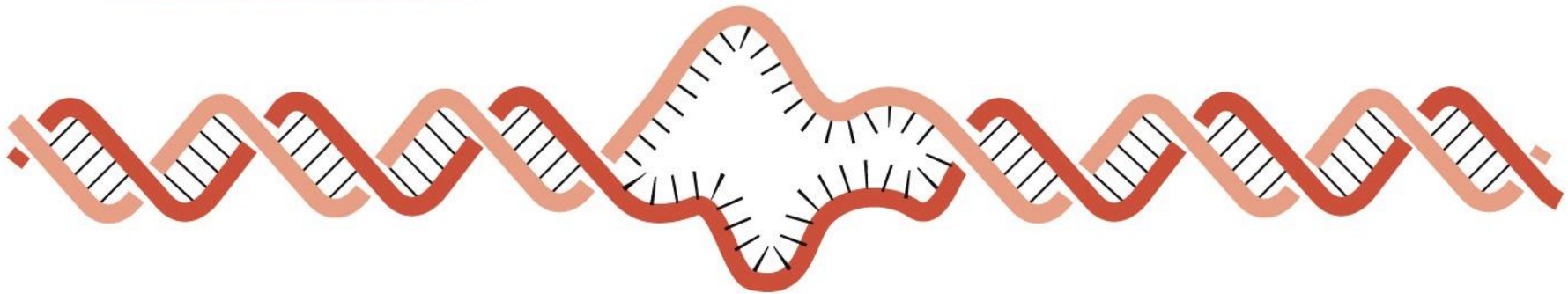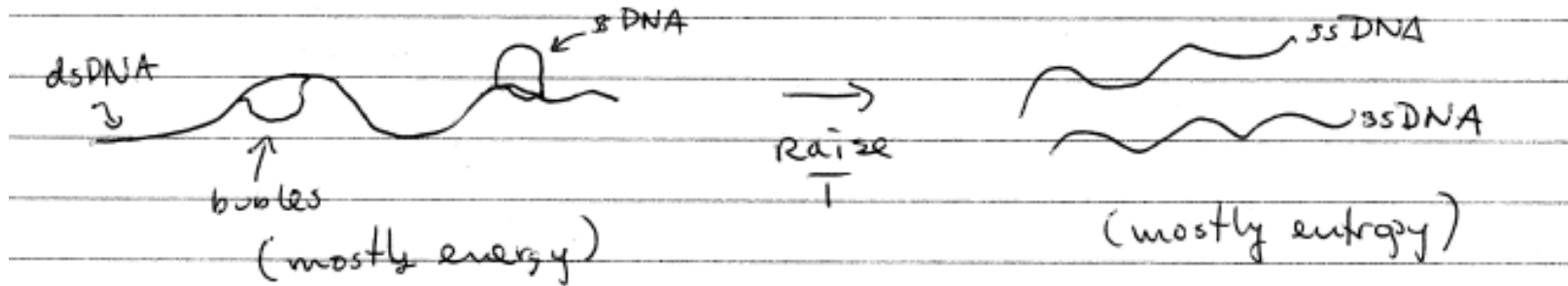

Figure 8.18a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

DNA melting is exploited in doing PCR – cycle of opening and closing DNA by changing the temperature

Q: At a given temperature, what is the typical bubble size in basepairs?

ssDNA

dsDNA

bubbles
(mostly energy)

Raise
T

ss DNA

ss DNA

(mostly entropy)

- The free energy cost (entropic cost) of bubbles can be modelled using a random walk model for the single stranded DNA.

- Can use model to calculate melting temperature

Random walk

n bp
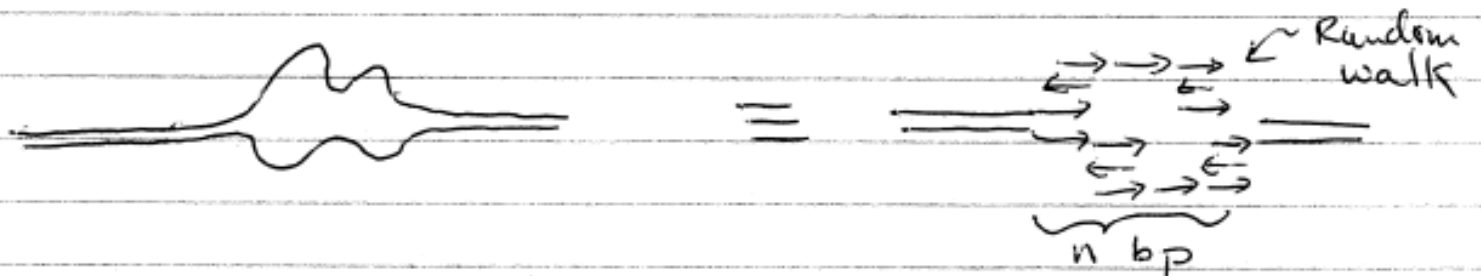
closed loop of a 1D random walk

Consider the free energy difference between the closed and bubble states

$$F_c$$

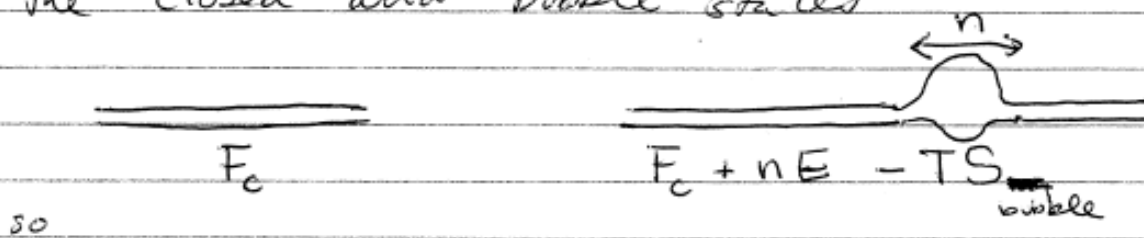$$F_c + nE - TS_{bubble}$$

so

$$\Delta F = (F_c + nE - TS) - F_c = nE - Tk_B \ln W_{bubble}$$

- here $E > 0$ is the energy associated with pairing a base pare.

- Now $W_{bubble} = \Omega_0(n) \cdot (N-n)$

  $\#$ of configs of 2 ss DNA of length n

  $\#$ of places to put a bubble of size n

  $= $ closed loop of length $2n$, returns to $x=0$

- need $n_r = n \rightarrow n_2 = n$ & $x = 0$

$$\Omega_o(n) = \frac{(2n)!}{n!\,(2n-n)!} = \frac{(2n)!}{n!\,n!}$$

use $\quad n! \cong \left(\frac{n}{e}\right)^n \sqrt{2\pi n} = \text{Stirling}$

$$\text{for } n \gg 1$$

Then

$$\Omega_o(n) = \frac{2^{2n}}{\sqrt{\pi n}}$$

so

$$\frac{\Delta F}{k_o T} = n \varepsilon - \cancel{1}\ln\left(\frac{2^{2n}}{\sqrt{\pi n}}(N-n)\right)$$

$$= n\underbrace{(\varepsilon - 2\ln 2)}_{E/k_B T} + \frac{1}{2}\ln n - \ln(N-n)$$

- At what value n_i^* does ΔF take a minimum?

- Solve for n^* from $\dfrac{d(\Delta F)}{dn} = 0$

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

Solution:

$$n^*_{\pm} = N\left(\frac{1 + \Delta\varepsilon \pm \sqrt{1 + 6\Delta\varepsilon + \Delta\varepsilon^2}}{\Delta\varepsilon} \right)$$

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

corresponds to $\dfrac{E}{k_B T} - 2\ln 2 > 0$

or

$k_B T < \dfrac{E}{2\ln 2}$ for no bubbles

- for $\Delta\varepsilon < 0$ or $k_B T > E/2\ln 2$ bubbles are possible

DNA melting and PCR: equilibrium bubble size
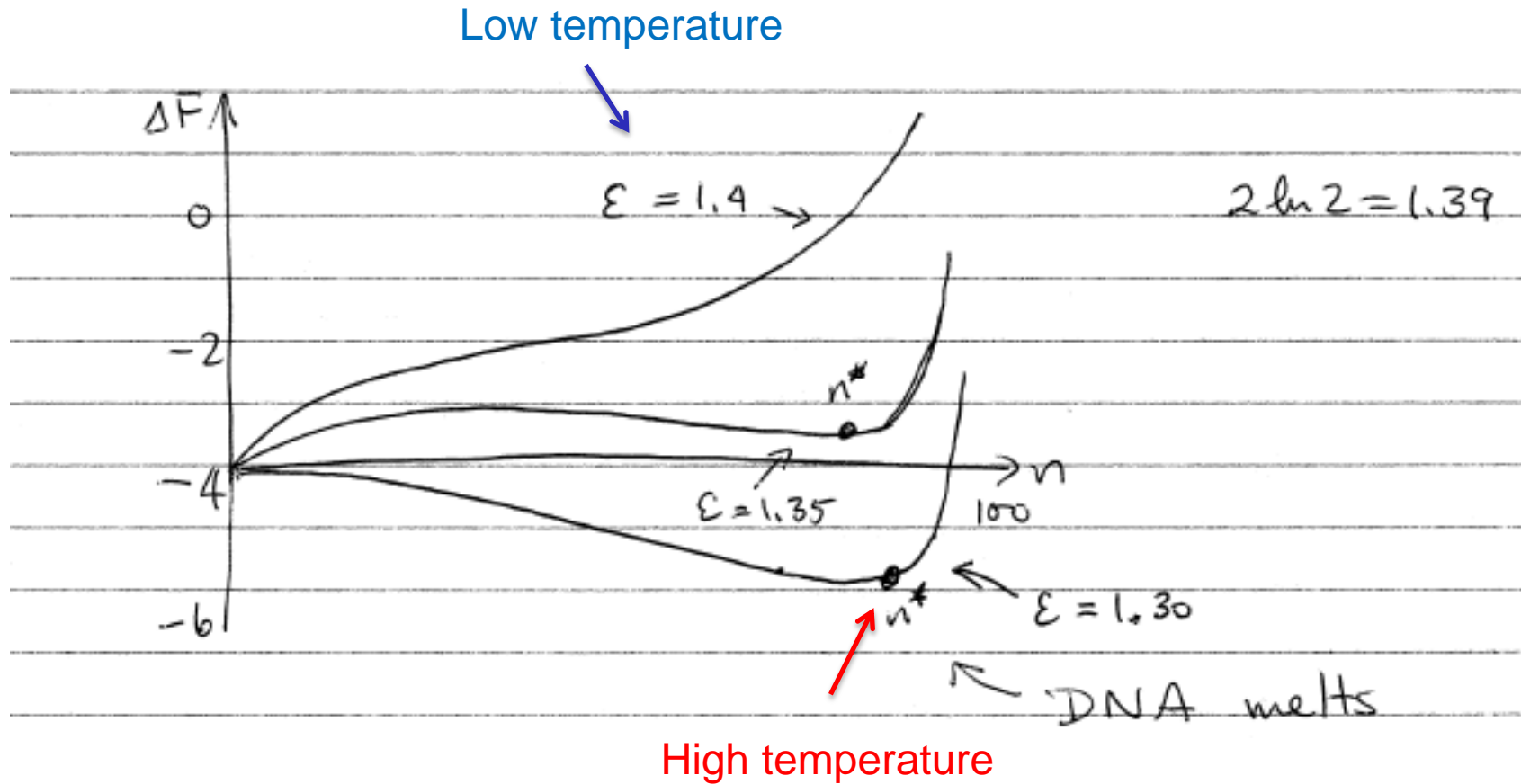
PCR works by cycling temperature over this melting transition

Atomic Force Microscope (AFM)

laser
beam

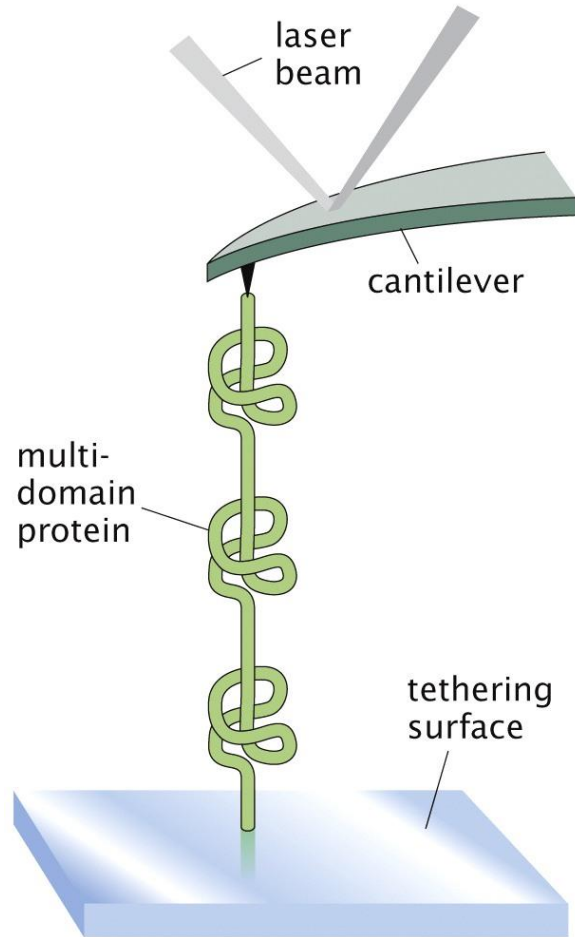cantilever

multi-
domain
protein

tethering
surface

Figure 8.21a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

If we apply a force to a random polymer to stretch it, what will it's 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
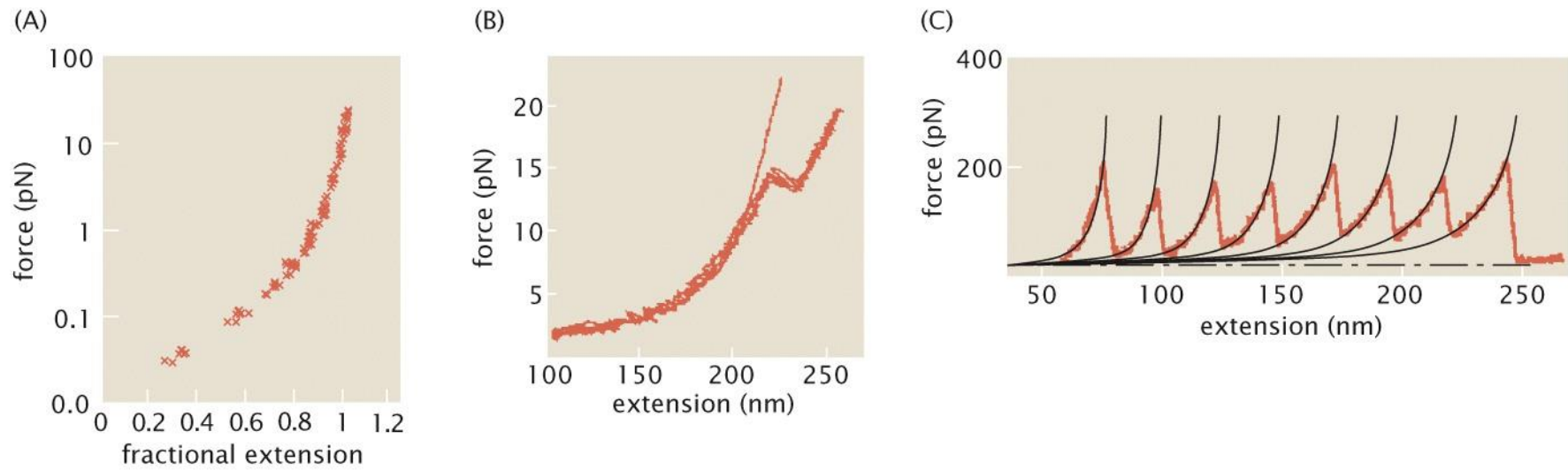
Pulling on a multidomain protein

(A)

(B)

(C)

Figure 8.22 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
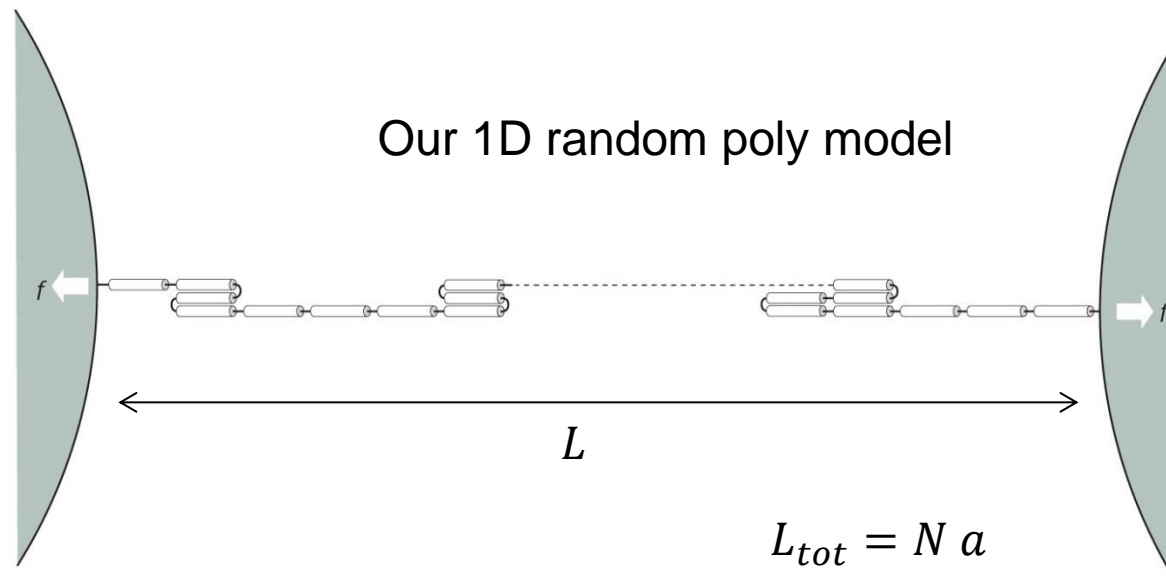
Our 1D random poly model

$$L_{tot} = N\,a$$

Figure 8.23 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Free energy of stretched freely jointed chain

- 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!

so
$$F(n_r) = -fn_r a + k_B T \left[n_r \ln n_r + (N-n_r) \ln (N-n_r) \right]$$

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

- Equilibrium $n_r^* \Rightarrow dF/dn_r = 0$

$$\Rightarrow -fa + k_b T \ln n_r - kT \ln n_L^{\nearrow N - n_r} = 0$$
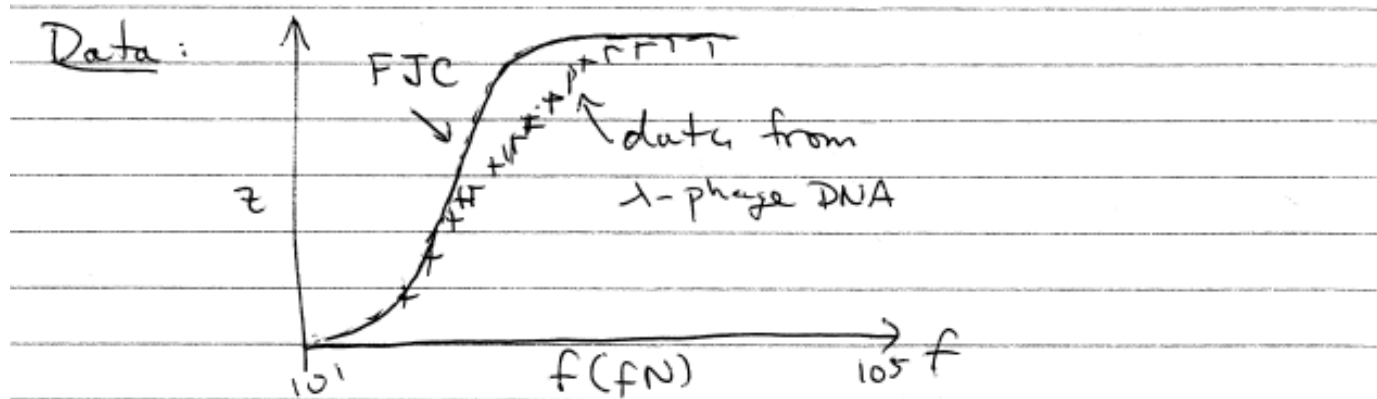
or $\quad \dfrac{n_R}{n_L} = e^{fa/kT}$

- % extension $\equiv z = \dfrac{\langle L \rangle}{\langle L_{tot} \rangle} = \dfrac{n_R - n_L}{n_R + n_L} = \dfrac{(n_R/n_L - 1)}{(n_R/n_L + 1)}$

so

$$\boxed{z = \tanh\left(\dfrac{fa}{kT}\right)}$$

Comparison to data:

Data plot: axes z (vertical) vs f (fN) (horizontal), from 10^1 to 10^5. Curves labeled FJC and "data from λ-phage DNA"

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

- Small force limit: $\dfrac{fa}{k_s T} \ll 1$

or $\quad 2f\xi \ll k_s T \quad$ or $\quad f \ll \dfrac{k_s T}{2\xi}$

for dsDNA : $\xi = 50\,nm \quad \rightarrow \quad f \ll 40\,fN$

for ss DNA : $\xi = 0.75\,nm \quad \rightarrow \quad f \ll 3\,pN$

Hook's law limit at low forces:

- Taylor expand $z = \tanh(x)$; $x = 2f\xi/k_BT$
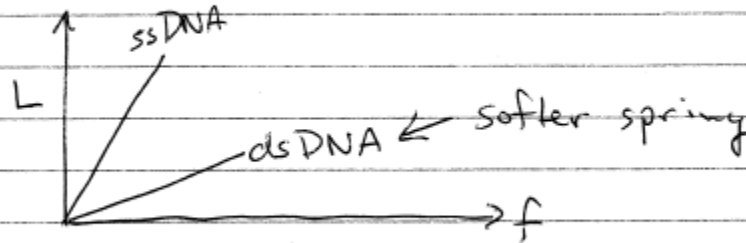
$$\Rightarrow \quad \langle L \rangle = \left[\frac{L_{tot} + 2\xi}{k_BT}\right] f \quad c.f \quad x = \frac{f}{k}$$

Hookian spring

- effective spring: $k = \frac{k_BT}{2L_{tot}\xi} \equiv$ entropic spring

for small forces

for λ-phage: $L_{tot} = 16\ \mu m$

ss DNA$_\lambda \rightsquigarrow k_{ss} = 160\ fN/\mu m$

ds DNA$_\lambda \rightarrow k_{ds} \simeq 2.3\ fN/\mu m$

- 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