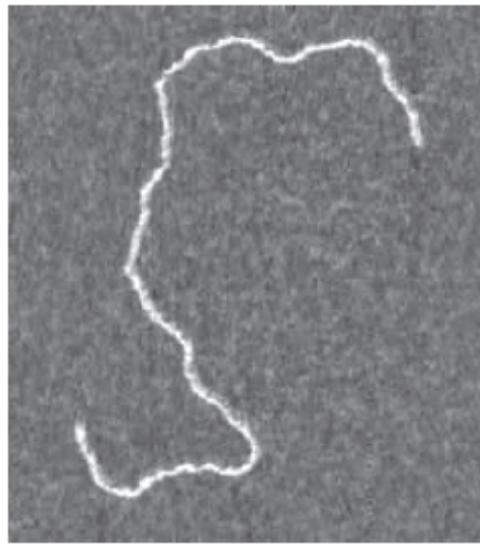


Today's class:

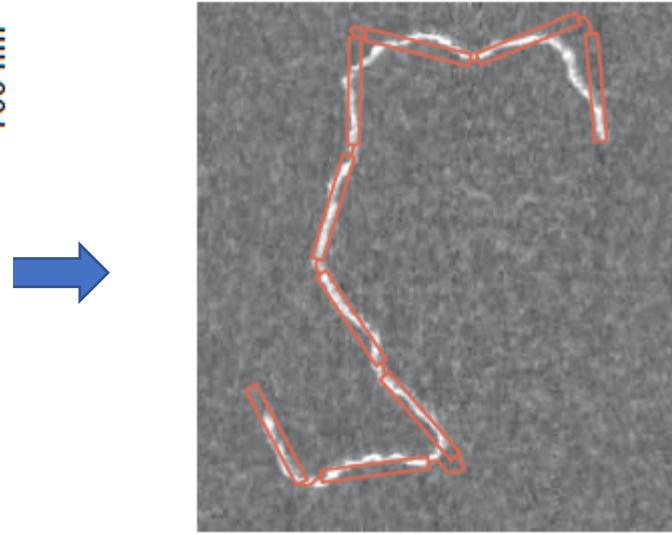
Random walk models of DNA and protein structures

*This lecture follows the parts of chapter 8 from the book
'Physical Biology of the Cell' by Philips, Konddev, Theriot and Garcia, 2nd Ed*

DNA structures can be represented as Random Walks



DNA on a surface captured by AFM
Wiggins et al Nature Nanotech 2006



DNA approximated by an array of randomly oriented segments

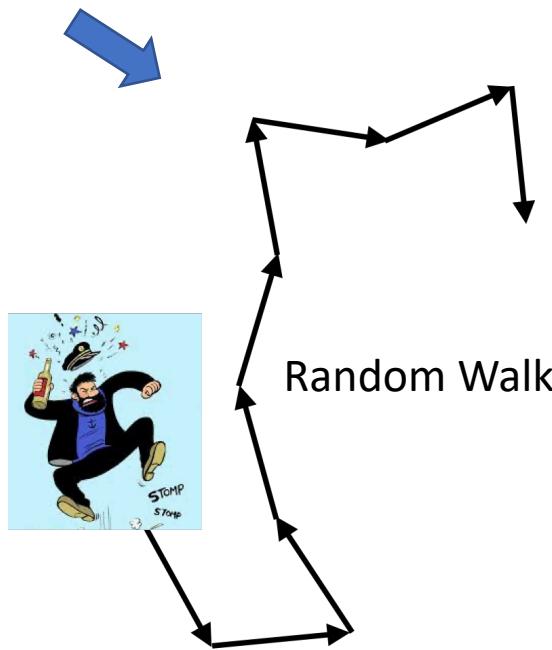
Each macromolecular configuration is a random walk along contour

Each step of random walk = one 'rigid' segment

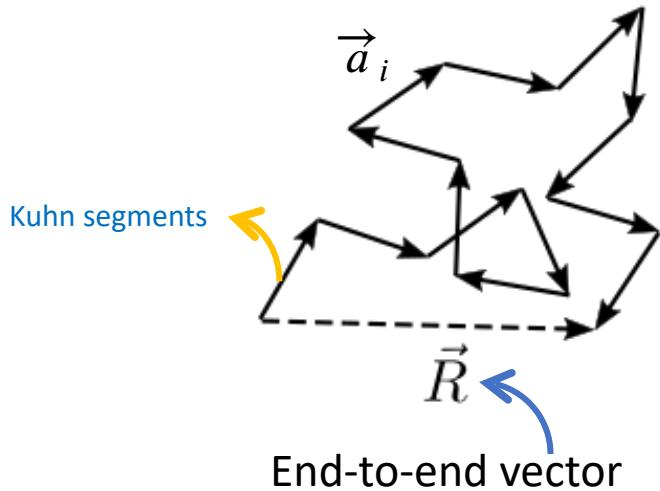
Independent steps, no dependence on history

For a 1D random walker, $p_R = p_L = 1/2$

For a chain of N segments, 2^N macromolecular configurations



The Freely Jointed Chain model



Non-interacting polymer of N Kuhn segments of length a
 Fully unfolded length, $L = Na$
 Some important results:

$$\langle \vec{R} \rangle = \left\langle \sum_{i=1}^N \vec{a}_i \right\rangle = \sum_{i=1}^N \langle \vec{a}_i \rangle = 0$$

$$\begin{aligned} \langle R^2 \rangle &= \left\langle \vec{R} \cdot \vec{R} \right\rangle = 0 \\ &= \left\langle \sum_{i,j}^N \vec{a}_i \cdot \vec{a}_j \right\rangle = \left\langle \sum_i^N a_i^2 \right\rangle + \left\langle \sum_{i \neq j}^N \vec{a}_i \cdot \vec{a}_j \right\rangle \end{aligned}$$

RMS end-to-end distance

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



$$\langle R^2 \rangle = Na^2$$

Volume occupied by the polymer

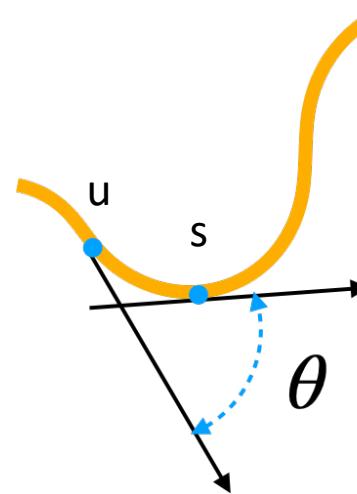
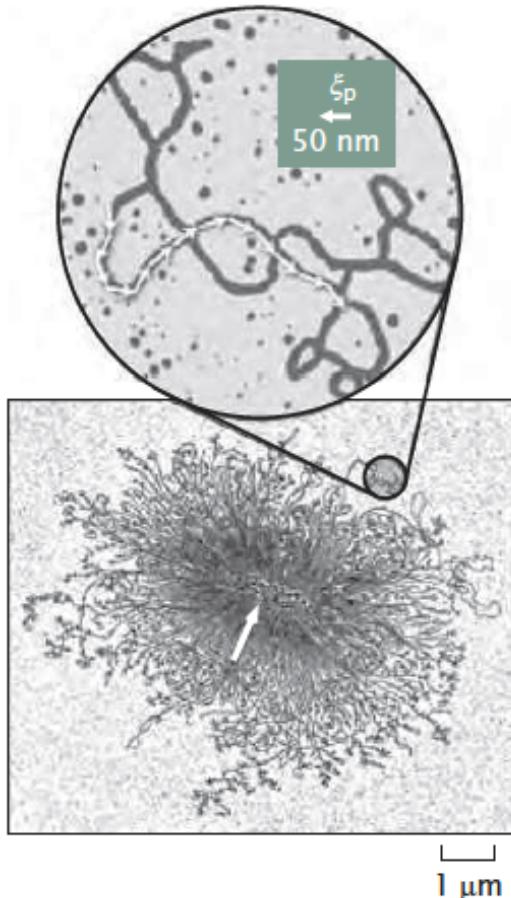
$$V_{poly} \propto \left(\sqrt{\langle R^2 \rangle} \right)^3 = a^3 N^{\frac{3}{2}} > \text{total volume of the unjoined monomers}$$

DNA is rigid up to some length and then bends

Persistence length
of a DNA

Length scale over which a DNA remains straight

How can one estimate the persistence length?



For any free polymer tangent-tangent correlation decays exponentially:

$$\langle \vec{t}(s) \vec{t}(u) \rangle = e^{-|s-u|/L_p}$$

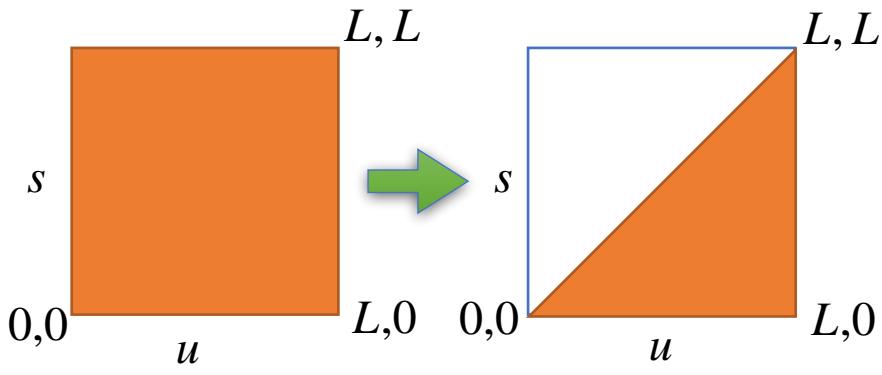
genome escaping out of a bacterial cell.

The decay length scale L_p defines the persistence length

How is persistent length related to Kuhn length?

For an elastic polymer of length L , end-to-end distance can be written as

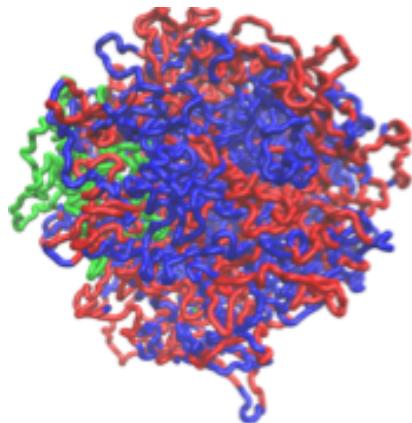
$$\begin{aligned} R &= \int_0^L ds \ t(s) \implies \langle R^2 \rangle = \left\langle \int_0^L ds \ \vec{t}(s) \cdot \int_0^L du \ \vec{t}(u) \right\rangle \\ &= 2 \int_0^L ds \cdot \int_s^L du \ \langle \vec{t}(s) \vec{t}(u) \rangle \\ &= 2 \int_0^L ds \cdot \int_s^L du \ e^{-(u-s)/L_p} \\ &= 2 \int_0^L ds \cdot \int_0^\infty dx \ e^{-x/L_p} \quad \text{for } x = u - s \quad \& L \gg L_p \\ &= 2LL_p \end{aligned}$$



From FJC model, $\langle R^2 \rangle = Na^2 = aL$, where a = Kuhn segment

Comparing the two results, $a = 2L_p$

Radius of gyration is a more precise measure of the size of the genome



Human Chromosome 1

Cheng et al eLife 2020

Radius of gyration is defined as the average distance of monomers from the center of mass of the polymer

$$\text{center of mass } \vec{R}_{CM} = \frac{1}{N} \sum_i R_i$$

\vec{R}_i = Position of i -th monomer

Then radius of gyration is given by

$$R_G^2 = \frac{1}{N} \sum_{i=1}^N \left(\vec{R}_i - \vec{R}_{CM} \right)^2$$

We can use the result $\langle R^2 \rangle = 2LL_p$

$$\text{to show eventually, } \sqrt{\langle R_G^2 \rangle} = \sqrt{\frac{LL_p}{3}}$$

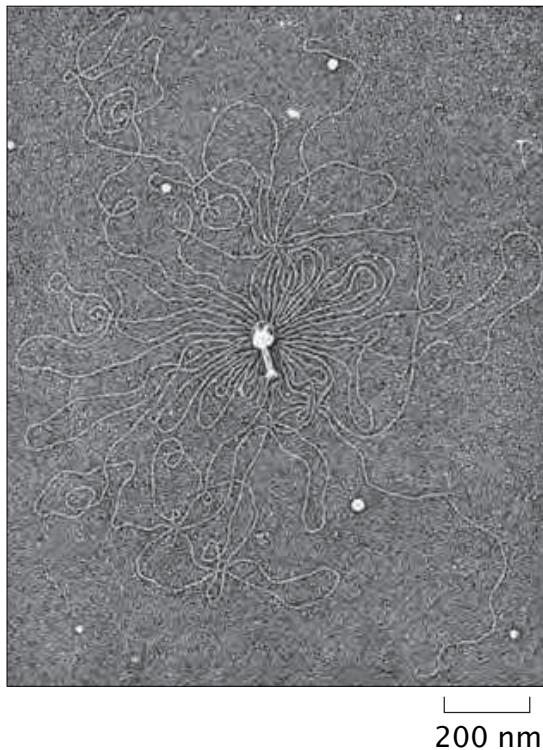
R.M.S radius of gyration $\propto L^{1/2}$

Radius of gyration and the size of the genome

assuming size of a single base-pair $\approx 0.34 \text{ nm}$ $\Rightarrow L \approx 0.34 \text{ nm} \times N_{bp}$

$$\Rightarrow \text{rms } R_G \approx \frac{1}{3} \sqrt{N_{bp} \times L_p} \text{ nm}$$

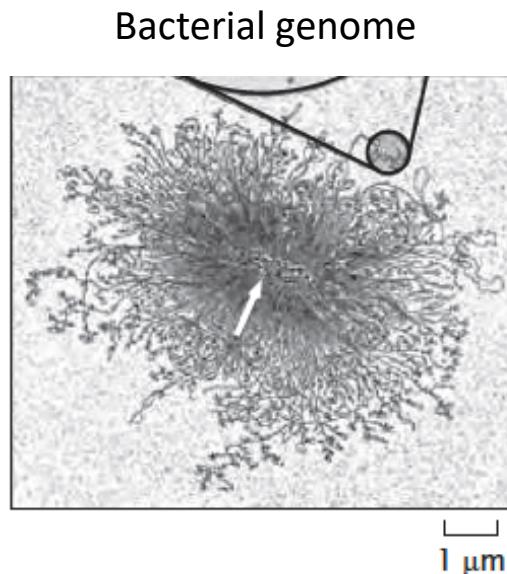
Electron microscopy image of a bacteriophage genome that has escaped its capsid.



Genome size $\sim 150 \text{ kb}$

$$\Rightarrow \text{rms } R_G = \frac{1}{3} \sqrt{150 \times 10^3 \times 50} \text{ nm} \approx 0.9 \mu\text{m}$$

This relation can be used to estimate the physical size of genome in solution



Genome size $\sim 5 \text{ Mb}$

$$\Rightarrow \text{rms } R_G \approx 5 \mu\text{m}$$

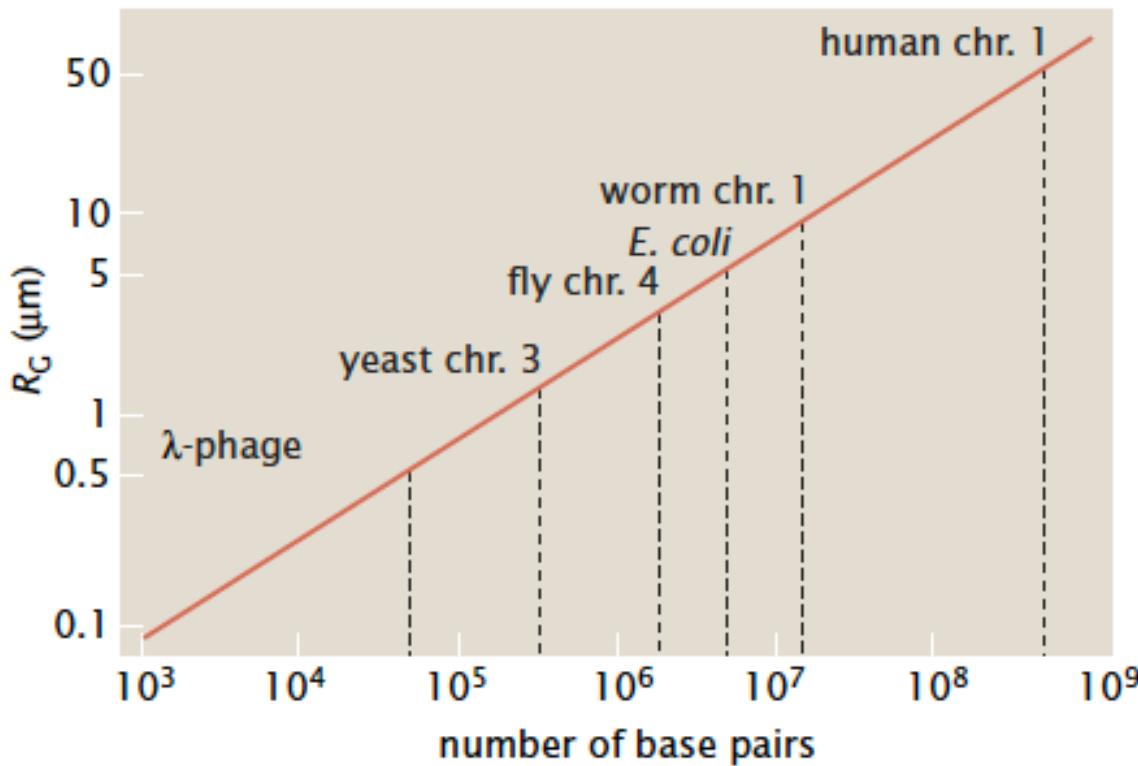
These are largely overestimates for actual genome sizes because both in the viral and the bacterial case the genome never takes such expanded configurations in the cell

Radius of gyration for different genomes

$$\text{rms } R_G \approx \frac{1}{3} \sqrt{N_{bp} \times L_p} \text{ nm}$$

This relation can be used to plot R_G vs. N_{bp}

Size of genomic DNA and specific chromosomes in solution estimated using random walk model



Because this measure of physical genome size is an overestimation, it thus can be a good upper bound for genome sizes

Chromatin Packing Density in nucleus

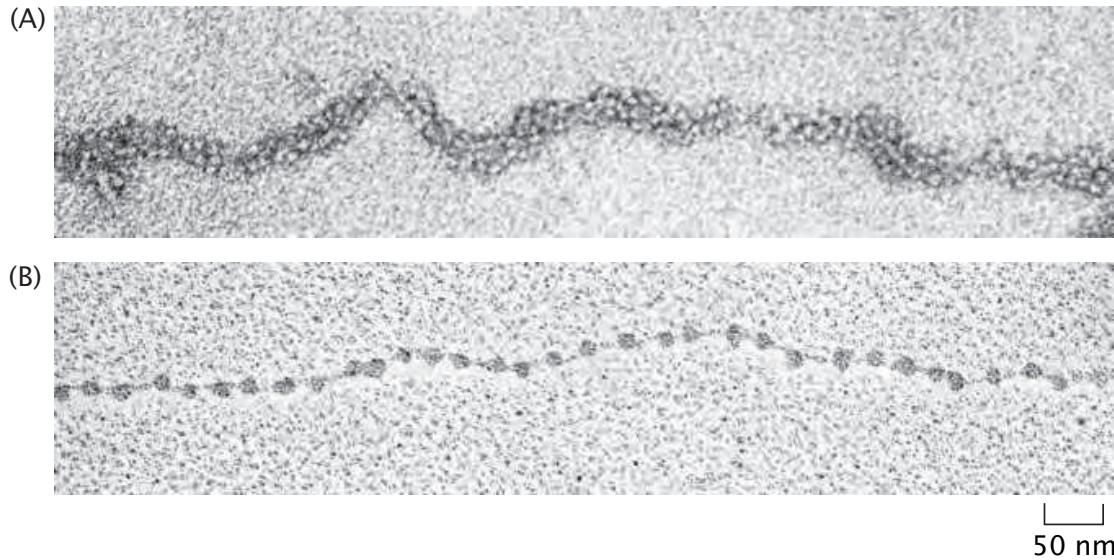


Figure 8.7: Electron microscopy images of chromatin. (A) Chromatin extracted from an interphase nucleus appears as a 30 nm thick fiber. (B) Stretching out a part of the chromatin reveals the “beads-on-a-string” structure of the 10 nm fiber, where each bead is an individual nucleosome. (A, courtesy of Barbara A. Hamkalo; B, courtesy of Victoria Foe.)

Here linear density of chromatin ν provides a measure of packaging

For 30 nm fiber model: $\nu = 100 \text{ bp/nm}$

For 10 nm fiber model: $\nu = ?$ From figure B, two nucleosomes per 50 nm

$$\Rightarrow \nu = \frac{2 \times 200 \text{ bp}}{50 \text{ nm}} = 8 \text{ bp/nm}$$

For comparison, for metaphase chromosomes, $\nu \approx 30,000 \text{ bp/nm} !$

Chromosome Packing in the Yeast Nucleus



Or



Some information

- Yeast cell has 16 chromosomes in the nucleus
- diameter of the interphase nucleus is about $\approx 2 \mu m$
- Total genome size $\approx 12 \text{ Mb}$

$$\text{Mean density } c = ? \quad c = \frac{12}{V_{nucleus}} = 3 \text{ Mb}/\mu\text{m}^3$$

What is the density for free chromatin? $c^* = N_{bp}/(4\pi R_G^3/3)$

Now we need to calculate R_G

How to get R_G for a free Yeast chromosome?

$$\text{rms } R_G = \sqrt{LL_p/3}$$

$$L = N_{bp}/\nu$$

$$N_{bp} = 12 \text{ Mb}/16 = 750 \text{ kb}$$

in vitro measured value of the persistence length for a 10 nm fiber is $L_p \approx 30 \text{ nm}$

For 10 nm fiber model: $\nu = 8 \text{ bp/nm}$

Finally, rms $R_G = 0.97 \mu\text{m}$

$$\implies L = 94 \mu\text{m}$$

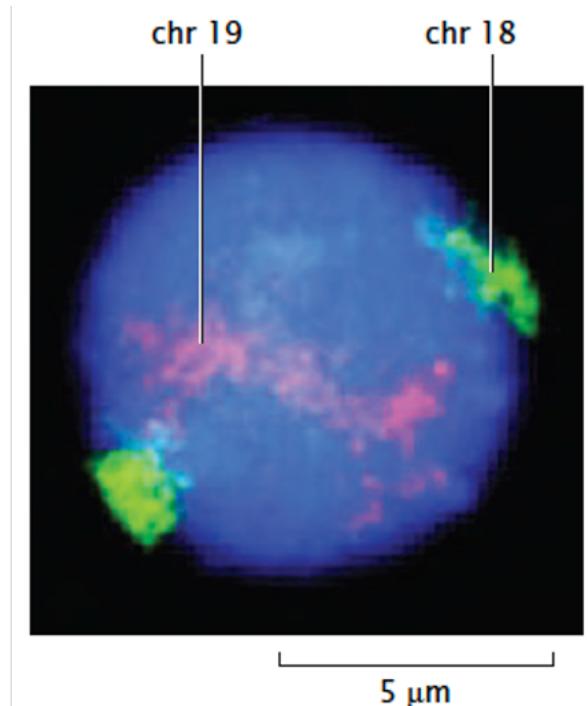
So, density of a free chromosome

$$\begin{aligned} c^* &= 750/(4\pi \times 0.97^3/3) \\ &\approx 200 \text{ kb}/\mu\text{m}^3 \end{aligned}$$

This value is about 15 times smaller than the nuclear chromatin density!!



But inside nucleus chromosomes appear as segregated!

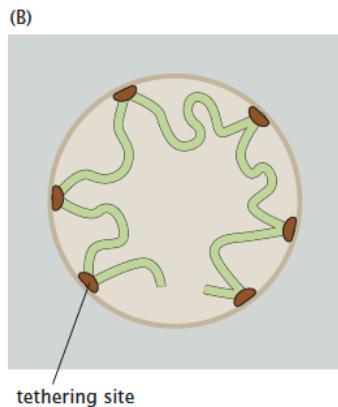
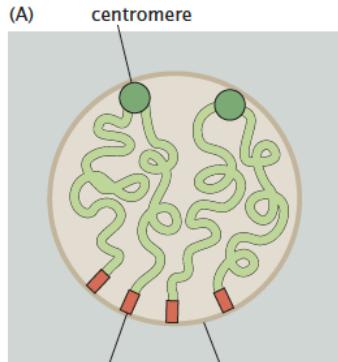


Fluorescently stained chromosomes 18 (green) and 19 (red) in the nucleus of a human cell.

What's the mechanism for segregation as polymer physics predicts entanglement?

We need something beyond a simple random walk model

Chromatin tethering on nuclear periphery can lead to segregation



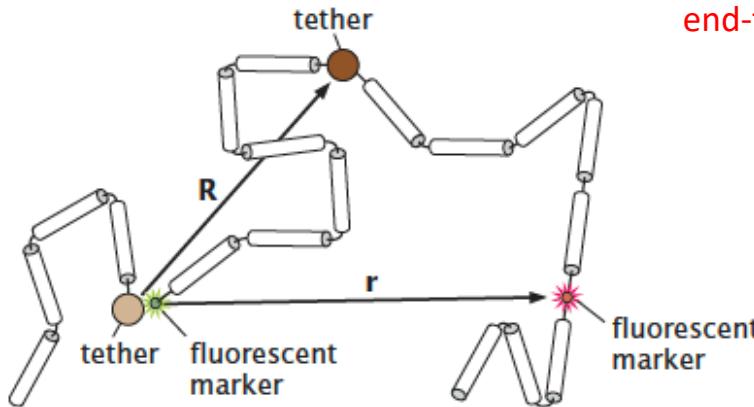
Distribution of end-to-end vector \vec{r} between two fluorescently tagged genomic loci with one tether present

$$P(\mathbf{r}) = \left(\frac{3}{2\pi N a^2} \right)^{3/2} e^{-3\mathbf{r}^2/2Na^2}$$

N = No. of segments between tether and marker

Based on the
3D form of the PDF of
end-to-end distance

Two tether scenario:



We can study this
using FISH microscopy

For two tethers, the
distribution becomes

$$P(\mathbf{r}) = \left(\frac{3}{2\pi N' a^2} \right)^{3/2} e^{-3(\mathbf{r}-\mathbf{R})^2/2N' a^2}$$

No. of segments between 2nd
tether and 2nd marker

These PDFs are not measurable exactly due to random orientations of cells

Chromatin tethering....contd

Due to random orientations of cells in experimental conditions measured distances are scalar not vectors. So, we use a converted form.

The end-to-end PDFs measured for the RWM:

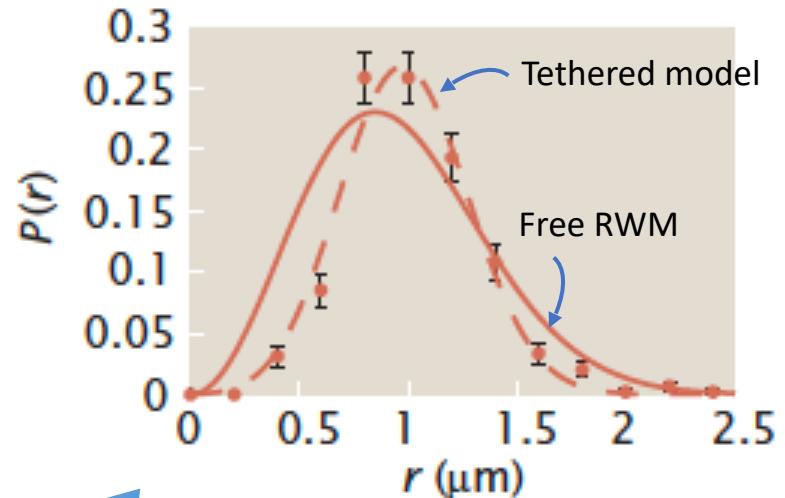
Free case

$$P(r) = \left(\frac{3}{2\pi N a^2} \right)^{3/2} 4\pi r^2 e^{-3r^2/2Na^2}$$

Tethered case

$$P(r) = \left(\frac{3}{2\pi N' a^2} \right)^{1/2} \frac{r}{R} \left(e^{-3(r-R)^2/2N' a^2} - e^{-3(r+R)^2/2N' a^2} \right)$$

The tethered model describes the data showing the robustness of the random walk model



Measurement of the distribution of distances between two tagged regions on yeast chromosome III

See also Avsaroglu et al Plos One 2014 for more details

$P(r)$ becomes a mathematical tool to detect tethering of chromosomes in cells!!

Justification of Flory theorem for genomic distances

Flory theorem: for dense polymer systems, distributions of distances between monomers are described by random walk statistics.

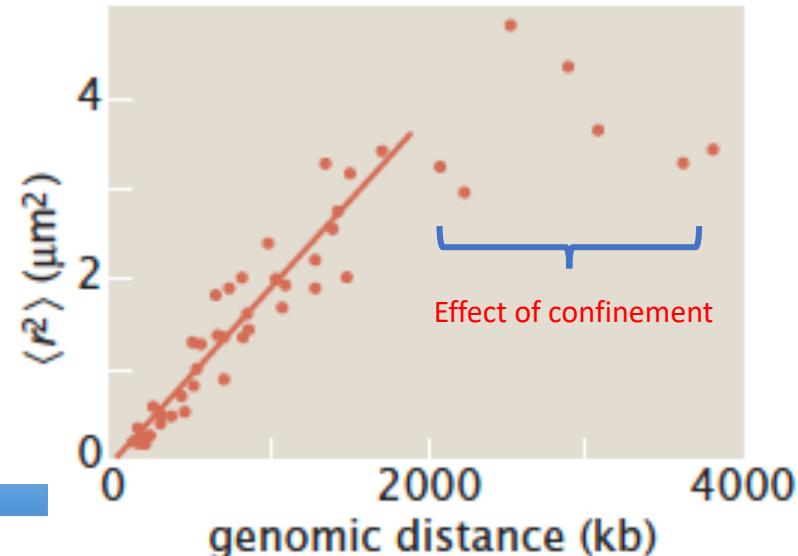
Let's see what we can achieve using this theorem

Distance between two labelled genomic loci $r = Na = N_{bp}/\nu$

where ν is the linear packing density of DNA
 N_{bp} is the genomic distance between the loci

According to random walk model $\langle r^2 \rangle = Na^2 = \frac{a}{\nu} N_{bp}$

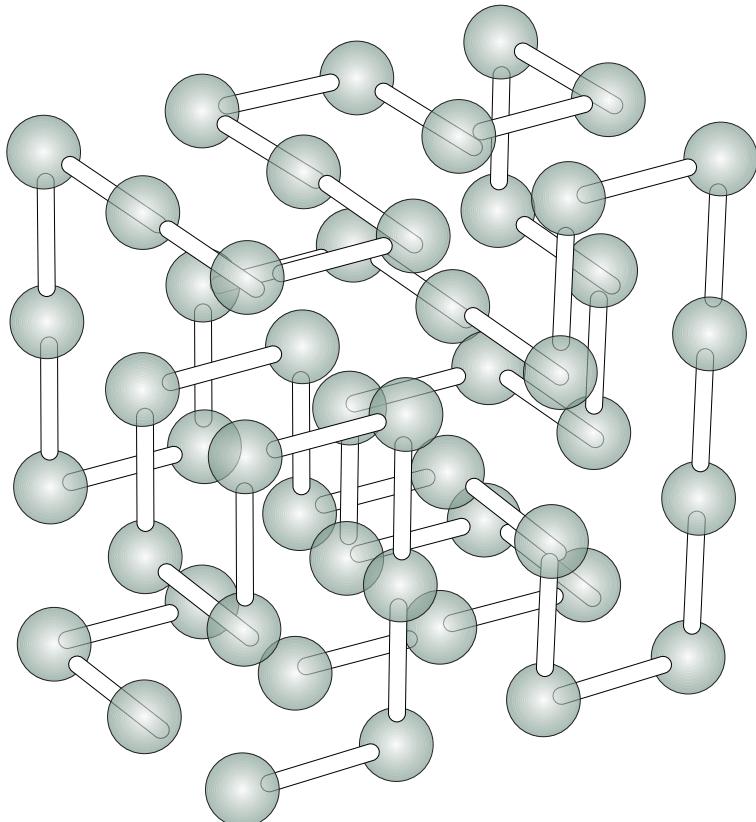
We get from the early slope, $\frac{a}{\nu} = 2 \text{ nm}^2/\text{bp}$



For human chromosome 4
G. van den Engh et al., Science 1992.

This exercise also provides a measure of chromatin length compaction by nuclear confinement

Proteins as random walks



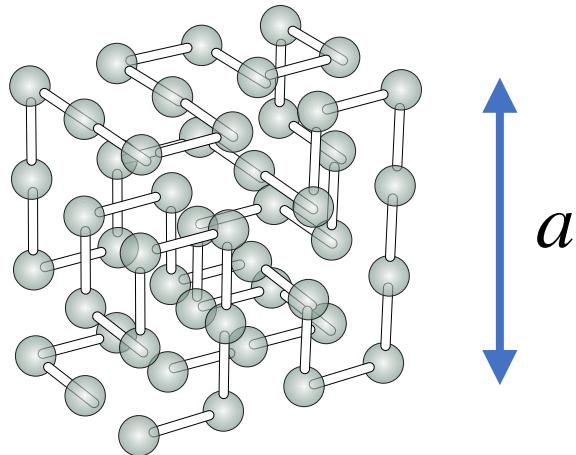
- Proteins can't be simple random walks
- Proteins are compact random walks
- They are self-avoiding random walks on a cubic lattice each site visited is occupied by an amino acid
- All sites are occupied on the lattice so entire surface sites are exposed to solvent

This very limited model of proteins can provide us some interesting estimates for physical size of real proteins

How physical size of proteins depends on their mass?

In the compact random walk model

$$\text{Linear size of protein } l_p \approx a \propto N_{\text{lattice}}^{1/3}$$



If each site = one amino acid residue

$$\text{Then, } l_p \propto N_{\text{residue}}^{1/3}$$

If amino acids have an average mass \bar{m}_{residue}

$$\text{Volume of the protein } \propto l_p^3 \propto N_{\text{residue}} \times \bar{m}_{\text{residue}}$$

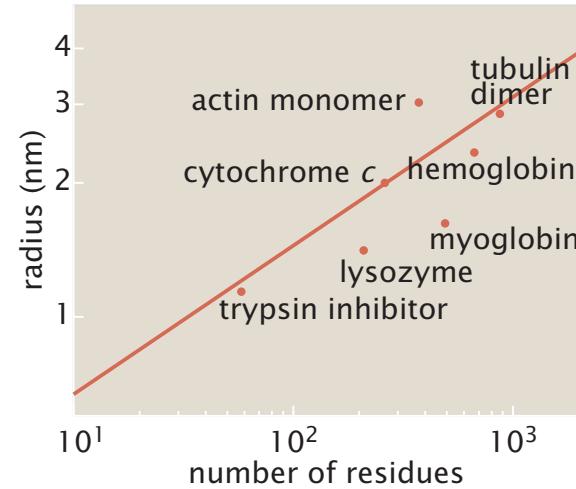


Figure 8.27: Scaling of protein size as a function of the number of amino acid residues. The line has a slope of $1/3$, corresponding to a space-filling packing.

However, for unfolded proteins the structure can be described as a simple random walk when the slope becomes $1/2$