Lecture 8:
Random Walks and the Structure of Macromolecules

Lecturer:
Brigita Urbanc
Office: 12-909
(E-mail: *brigita@drexel.edu*)

Course website:
www.physics.drexel.edu/~brigita/COURSES/BIOPHYS_2011-2012/

Description of the macromolecular structure as *random walks*

-motivation: calculate entropic cost of DNA packed into a cell; description of DNA stretching by optical tweezers

- atomic coordinates deposited on databases such as Protein Data Bank (pdb format of an ascii file with x,y,z coordinates of known protein structures)

$$(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N) \text{ with } \mathbf{r}_i = (x_i, y_i, z_i)$$

- data derived based on X-ray crystallography or NMR

- statistical measure of the size of a macromolecule: Radius of Gyration R_G

- example of DNA characterization $\mathbf{r}(s)$ where s is the distance along the contour of the molecule

Random walk model of a polymer:
series of rigid rods (the Kuhn segments) connected by flexible hinges

lattice model in 1D **lattice model in 3D**

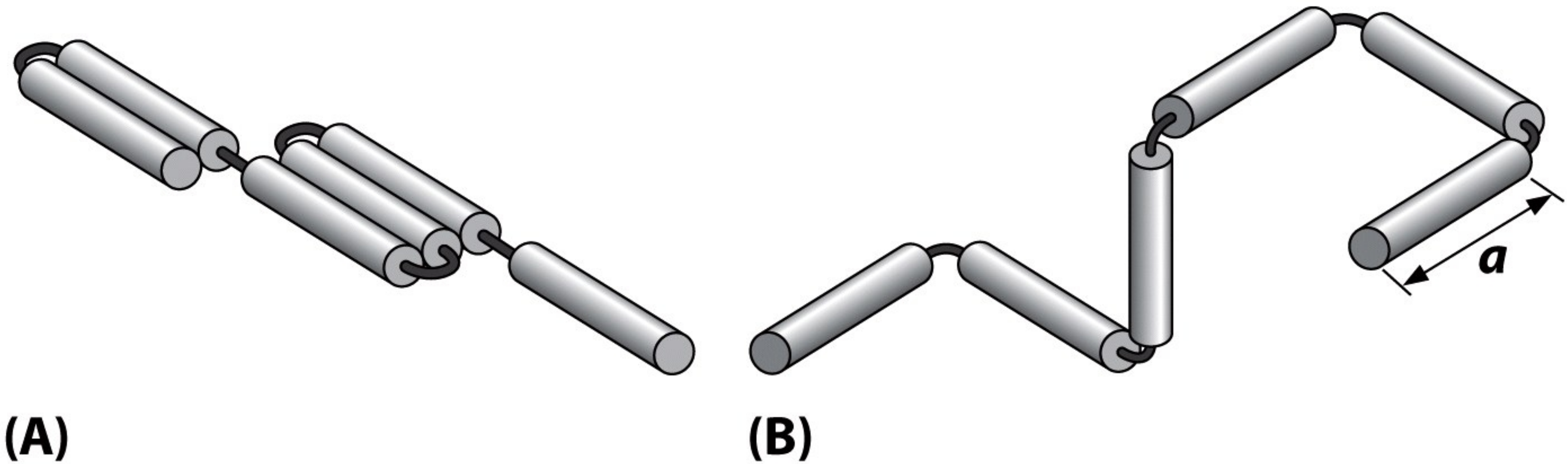

Figure 8.1 Physical Biology of the Cell (© Garland Science 2009)

DNA molecule as a random walk:
Every macromolecular configuration is equally probable

AFM image

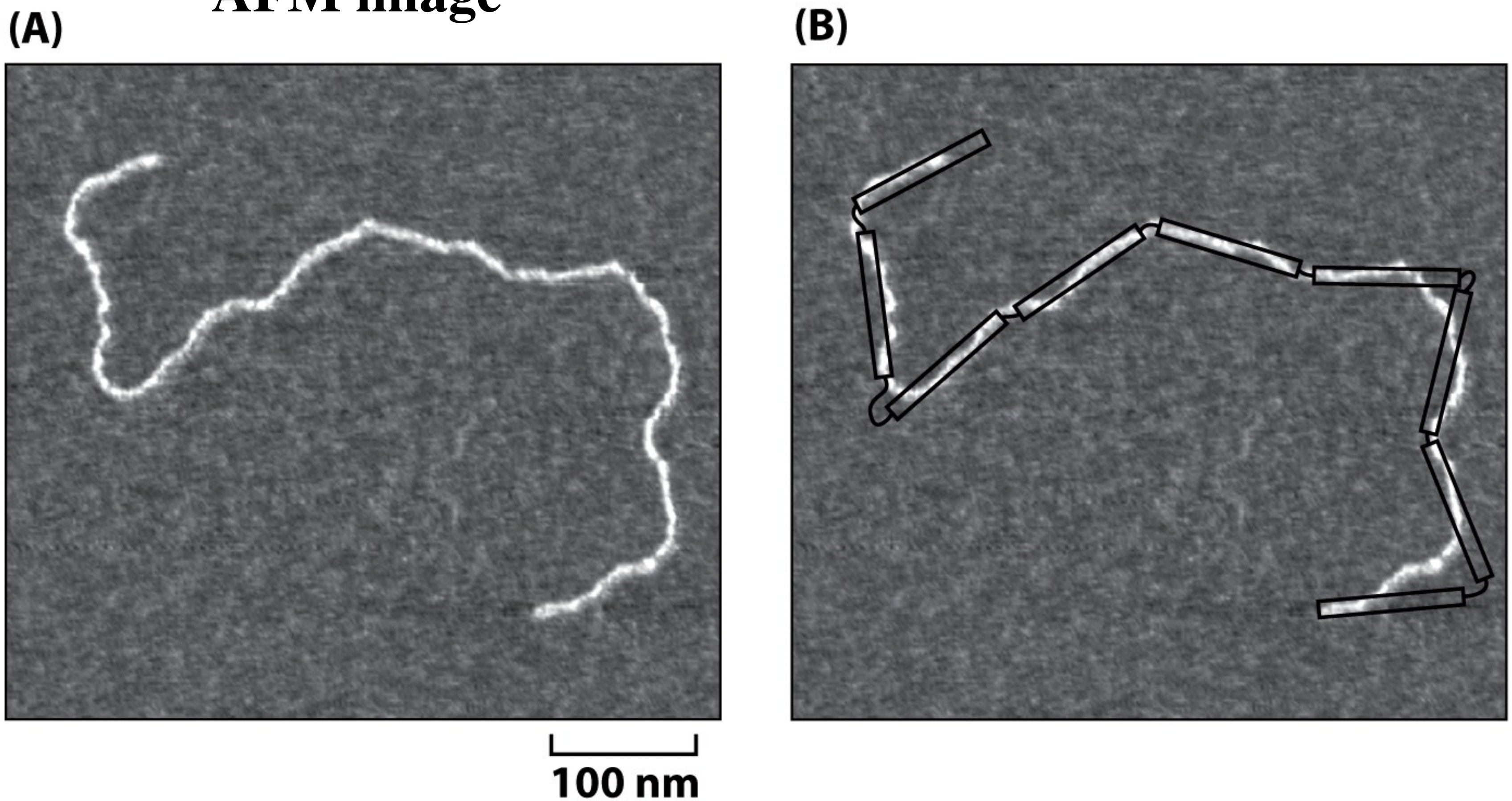

Figure 8.2 Physical Biology of the Cell (© Garland Science 2009)

What is a random walk?

- 1D random walker: $p_R = p_L = ½$ at every step, independently

- for N random walk steps, there are 2^N possible macromolecular configurations

- the mean distance of the walker from its point of departure:

$$\langle R \rangle = \langle \sum_i x_i \rangle = \sum_i \langle x_i \rangle = 0$$

- the variance of the probability distribution:

$$\langle R^2 \rangle = \langle \sum_i \sum_j x_i x_j \rangle = \sum_i \langle x_i^2 \rangle + \sum_{i \neq j} \langle x_i x_j \rangle$$

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

- thus the size of the area covered by the random walk is:

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

Derivation Based on the Probability Theory:

- a macroscopic state:
  many microscopic states

- N steps in 1D random walk:
  $N = n_R + n_L$, each with
  a probability of ½

- a probability associated with
  each microstate of N steps
  is: $(1/2)^N$

- multiplicity W associated with
  n_R steps is:

$$W = N!/[n_R!\,(N - n_R)!]$$

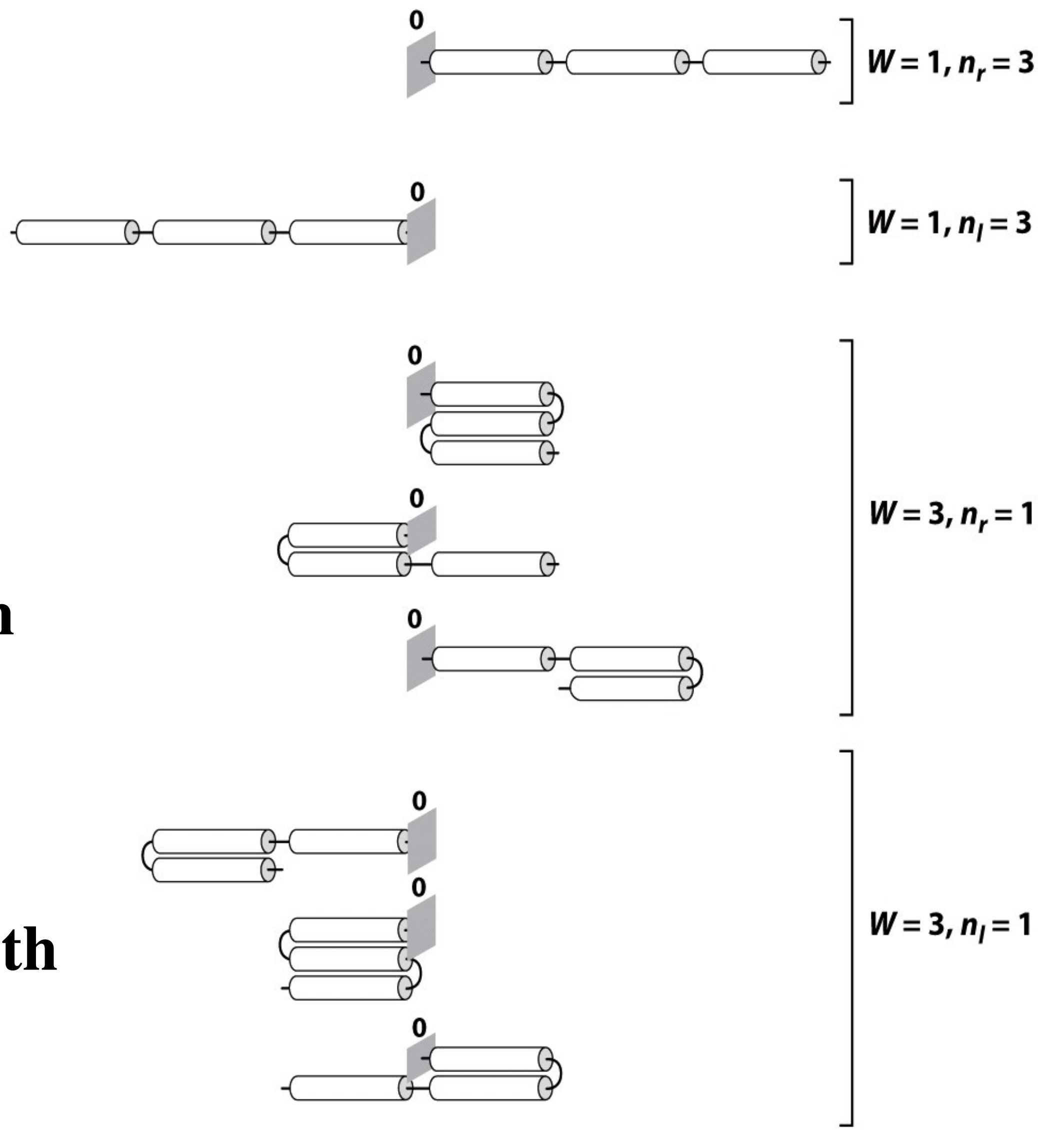

Figure 8.3 Physical Biology of the Cell (© Garland Science 2009)

Probability of an overall departure n_R from the origin:

$$p(n_R; N) = N!/[n_R! (N- n_R)!] (1/2)^N$$

This distribution is normalized:

$$\sum_{nR} p(n_R; N) = 1, \quad n_R \in \{0,N\}$$

Probability distribution for the end-to-end distance:

$$R = a (n_R - n_L) = a (2n_R - N) ; n_R = N/2 + R/(2a)$$

such that

$$p(R; N) = (1/2)^N N!/\{[N/2 + R/(2a)]! [N/2 – R/(2a)]!\}$$

which takes the form of a Gaussian distribution for $R \ll Na$:

$$p(R; N) = 2/\sqrt{2\pi N} \exp[-R^2/(2Na^2)]$$

To obtain the probability distribution function for the end-to-end distance of a freely jointed chain P(R; N), p(R; N) needs to be divided by the number of integer R values per unit length (=2a):

$$P(R; N) = 2/\sqrt{2\pi N a^2}\ \exp[-R^2/(2Na^2)]$$

Note that $\langle R \rangle = 0$ and $\langle R^2 \rangle = Na^2$, independent of the space Dimension. This can be used to derive P(R; N) in 3D space:

$$P(R; N) = A \exp(-\kappa R^2);$$

$$A = [3/(2\pi N a^2)]^{3/2};\ \ \kappa = 3/(2Na^2)$$

Use normalization condition in 3D and calculate the variance to determine A and κ (integration in 3D).

End-to-end distance probability distribution for 1D random walk: Binomial (dots) versus Gaussian (curve) distribution for N=100

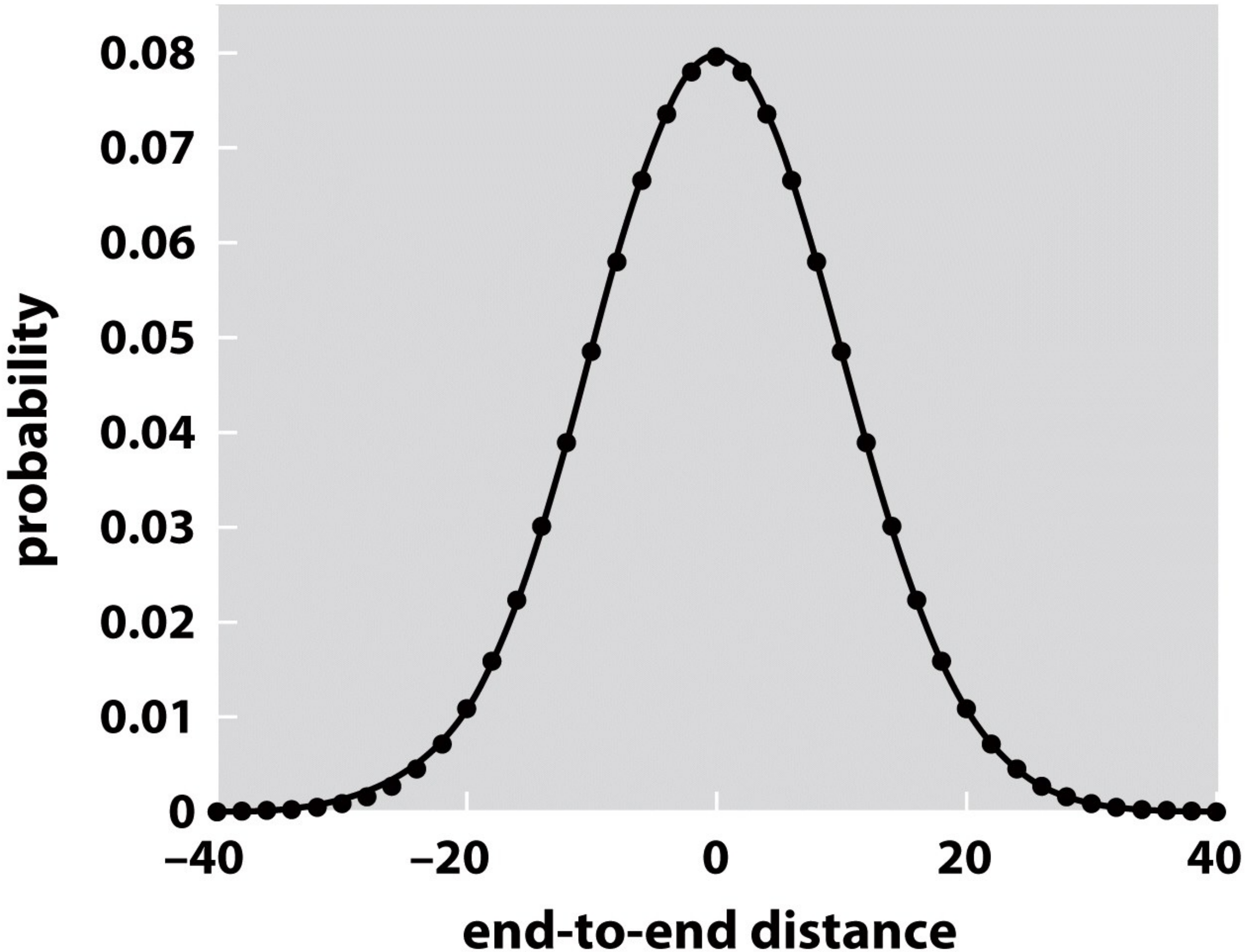

Figure 8.4 Physical Biology of the Cell (© Garland Science 2009)

Persistence Length versus the Kuhn Length

- *persistence length*: the length scale ξ_P over which the polymer remains approximately straight
- *the Kuhn length*: the length of a step a in the random walk model

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

- persistence length ξ_P can be calculated from the unit tangent vector $\mathbf{t}(s)$ where s is a distance along the polymer:

$$\langle \mathbf{t}(s)\, \mathbf{t}(u) \rangle = \exp(-|s-u|/\xi_P)$$

- to find the relationship between a and ξ_P we use $\mathbf{R} = \int_0^L ds\, \mathbf{t}(s)$:

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

$$2 \int_0^L ds \int_s^L du\, \exp[-(s-u)/\xi_P] = 2 \int_0^L ds \int_0^\infty dx\, \exp(-x/\xi_P) = 2L\xi_P$$

$$a = 2\xi_P$$

Size of genomic DNA in solution

$\langle R^2 \rangle = 2L\xi_P$

$\langle R_G^{\,2} \rangle = 1/3\ L\xi_P$

(use definition)

$L = 0.34\ \text{nm}\ N_{BP}$

$N_{BP\,\ldots}$ # of basepairs

$R_G = 1/3\ \sqrt{N_{BP}\,\xi_P}\ \text{nm}$

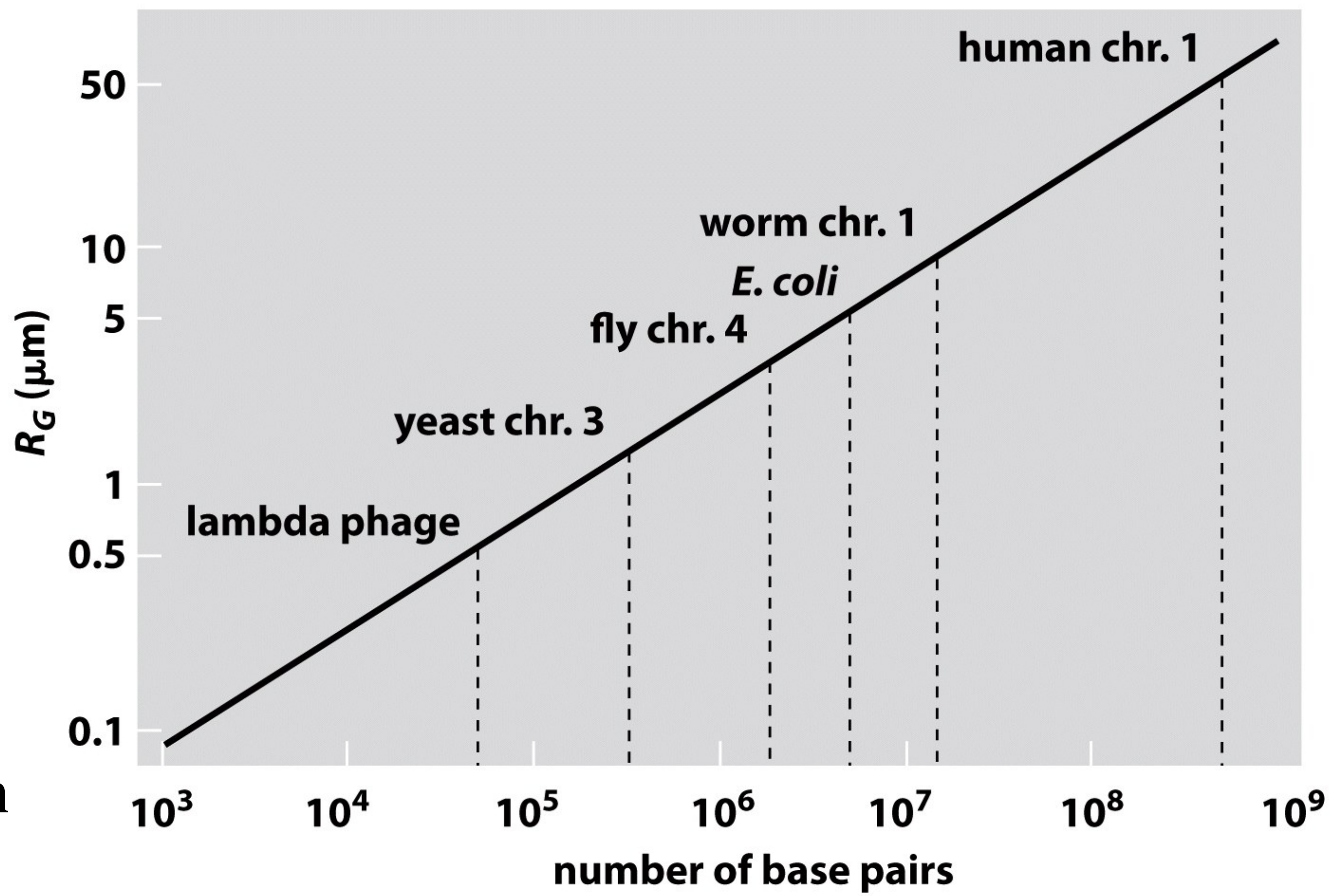

Figure 8.5 Physical Biology of the Cell (© Garland Science 2009)