Lectures 9 and 10:
Random Walks and the Structure of Macromolecules (contd.)

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

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

Bacterial genome that escaped the bacterial cell

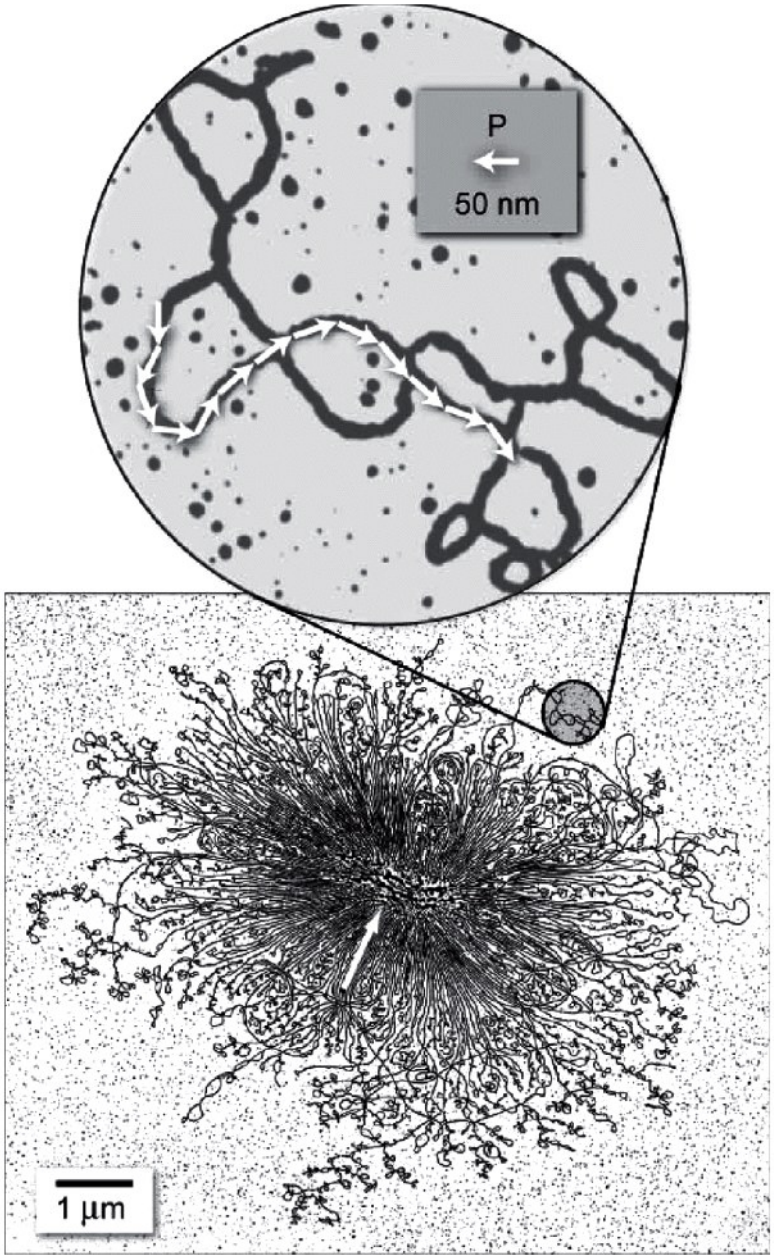


Figure 8.6 Physical Biology of the Cell (© Garland Science 2009)

DNA organization:

- **packed into chromatin fibers**
- **a unit of chromatin is nucleosome**
- **characterized by various packing densities**
- **linear density of chromatin ν [bp/nm]**

30 nm fiber
(ν=100 bp/nm)

10 nm fiber
(ν=8 bp/nm)

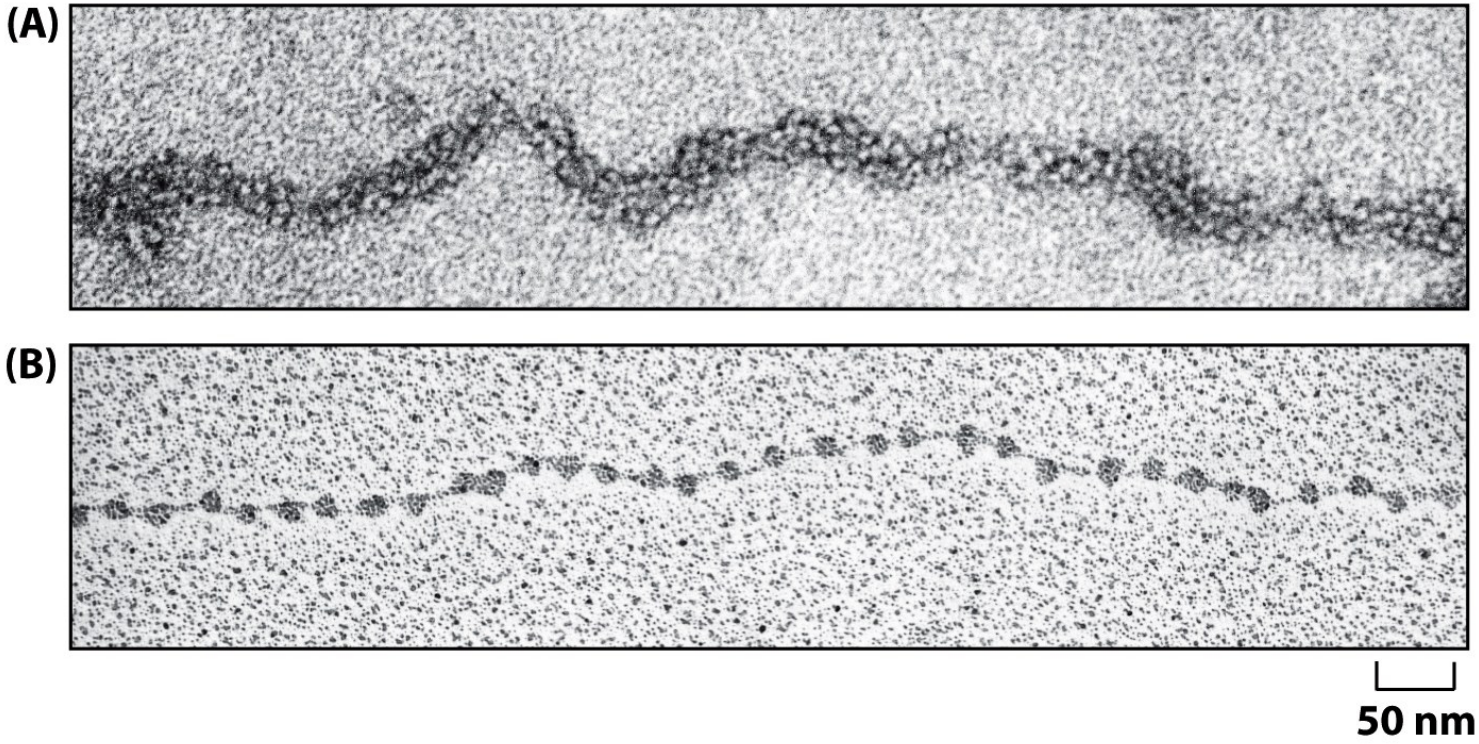


Figure 8.7 Physical Biology of the Cell (© Garland Science 2009)

Chromosomes 18 and 19 in a nucleus of a human cell (observed by fluorescence spectroscopy)

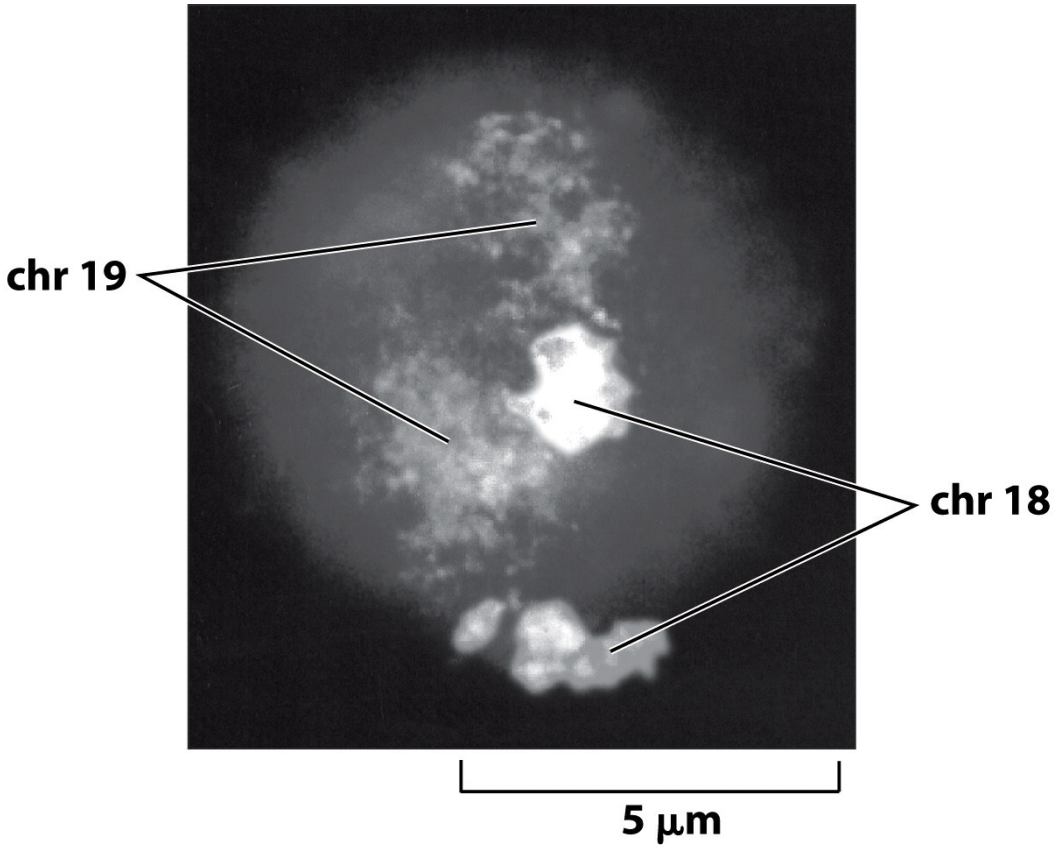


Figure 8.8 Physical Biology of the Cell (© Garland Science 2009)

- existence of chromosome territories
- chromatin density: a dense polymer system
- free polymers in a dense system inter-penetrate
- why is chromatin packed and localized?
- existence of *tethering*

Chromosome Packing in the Yeast Nucleus: Spaghetti vs. Meatballs

- yeast: 16 chromosomes in its nucleus with a diameter of 2 μm
- chromosome (= DNA molecule) size: 230 kb to 1,500 kb
  total genome size: 12 Mb (mega base pairs)
- thus a mean density $c = 12\ \text{Mb}/V_N$; $V_N = 4/3\pi \times 1\mu\text{m}^3$
  $c = 3\ \text{Mb}/\mu\text{m}^3$ … chromosome density inside a nucleus

- $c^* = N_{BP} / V_F$; $V_F = 4/3\pi \times R_G^3$ … chromosome solution density
- estimate R_G:based on the random walk model of a polymer:
  - +the length of a polymer 12 Mb/16 = 750 kb
  - +packing density of 8 bp/nm; L = 750 kb/8bp nm = 94 μm
  - +an *in vitro* measured ξ_P = 30 nm (for 10 nm fiber)
  - +$R_G = (L\xi_P/3)^{1/2}$ = 0.97 μm and $c^* = 200\ \text{kb}/\mu\text{m}^3$ …(~0.1 c)

Chromosomes are tethered at different locations inside the nucleus

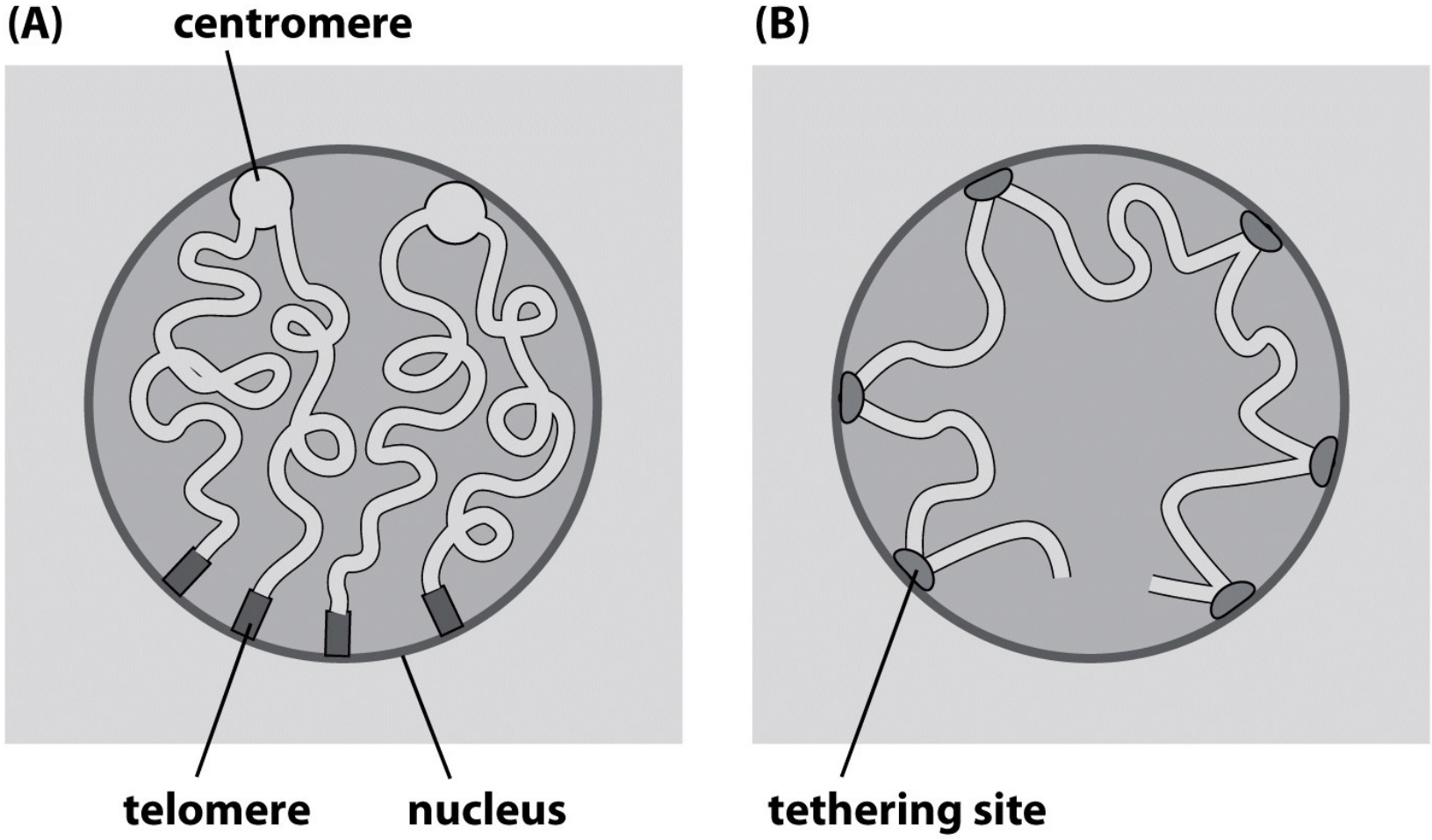
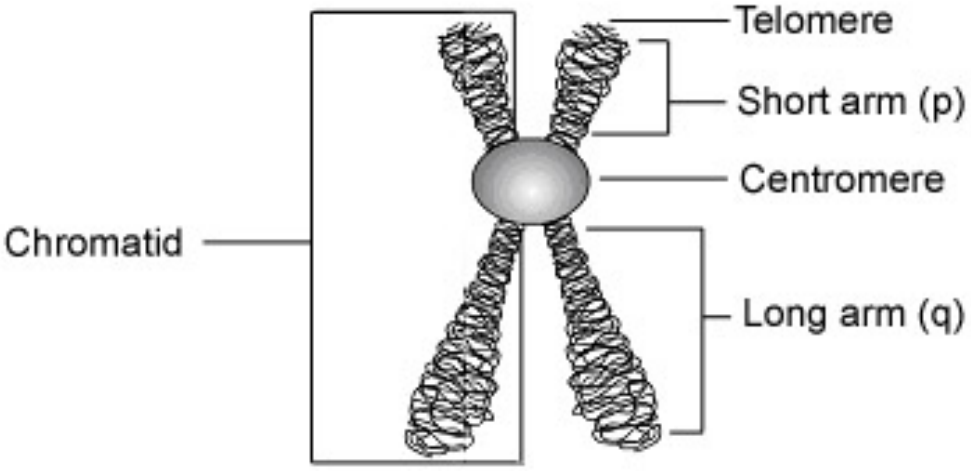
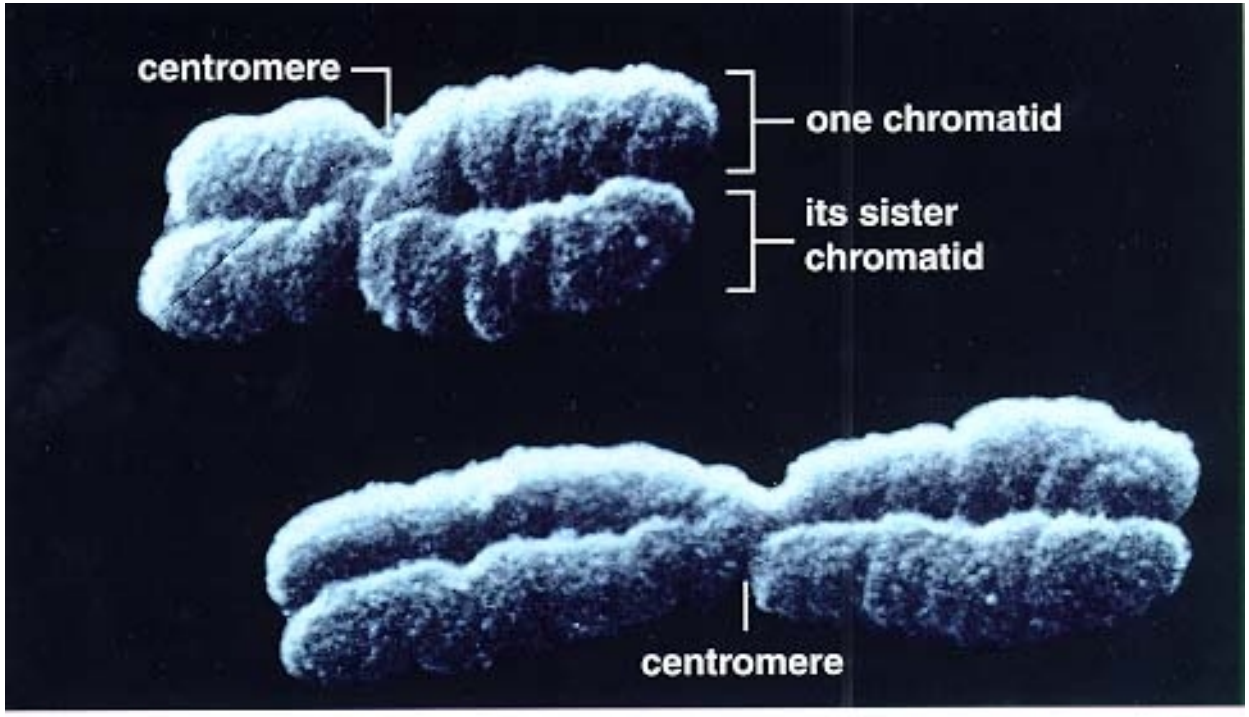


Figure 8.9 Physical Biology of the Cell (© Garland Science 2009)


Telomere
Short arm (p)
Centromere
Chromatid
Long arm (q)


centromere
one chromatid
its sister
chromatid
centromere

Experimental trick to examine the chromosome geography: Bind to DNA sites fluorescently labeled proteins

R – known genomic distance (number of base pairs) between two tethers that are fixed in space

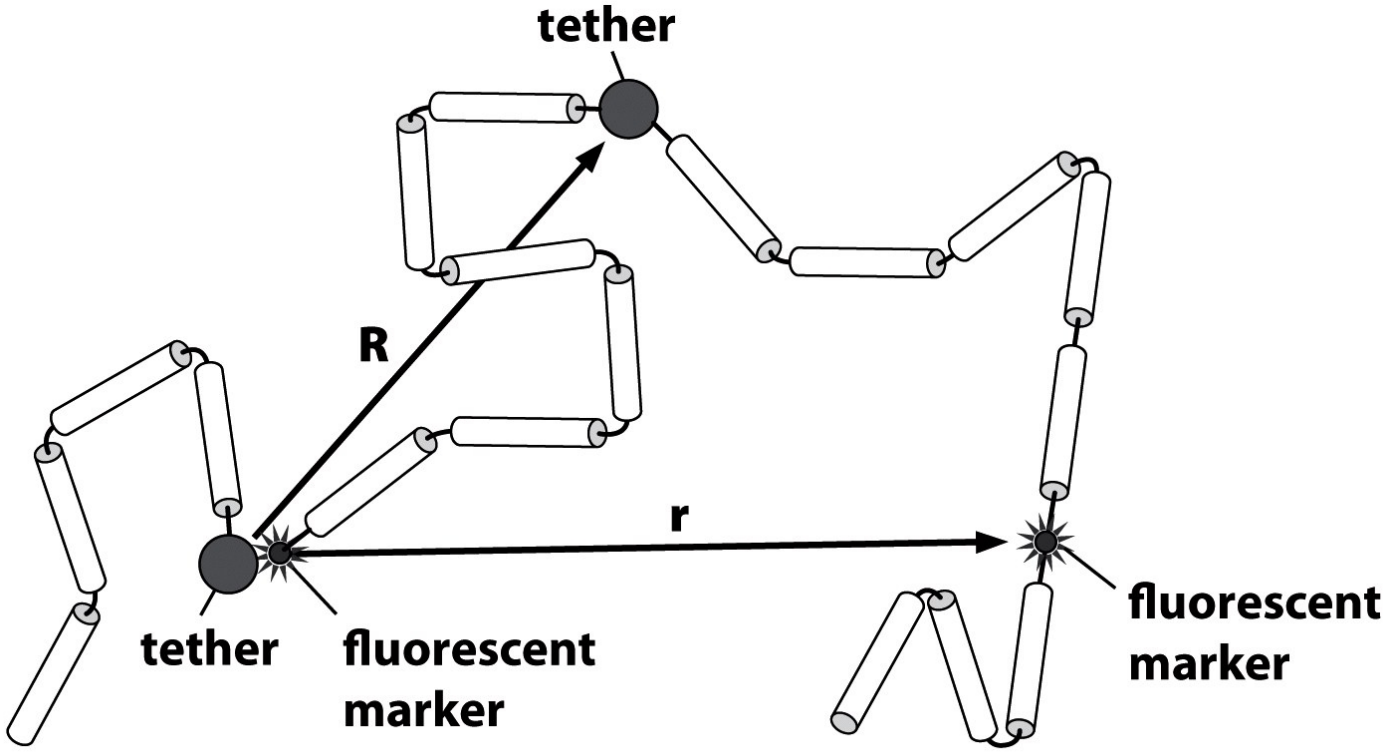


Figure 8.10 Physical Biology of the Cell (© Garland Science 2009)

None or only one tethering site, a random walk model of chromatin predicts *a Gaussian distribution*:

$$P(r) = C \exp(-r^2/\sigma^2)$$

$$C = [3/(2\pi N a^2)]^{3/2} \;\&\; \sigma^2 = 2Na^2/3 \;\&\; a = 2\xi_P$$

N … total number of Kuhn's segments (of length a);

L = Na … length of the polymer; r … a 3D vector

For two tethering sites (and two fluorescent markers) leads to *a displaced Gaussian distribution*:

$$P(r) = C' \exp[-(r-R)^2/\sigma'^2]$$

$$C' = [3/(2\pi N' a^2)]^{3/2} \;\&\; \sigma'^2 = 2N'a^2/3 \;\&\; a = 2\xi_P$$

N' … the number of Kuhn's segments between the 2nd tether and 2nd fluorescent marker

R and r … 3D vectors

Can the chromatin which is densely packed in a nucleus of a cell be approximated by a random walk model?

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

$Na = N_{BP}/\nu$

N_{BP} ... genomic distance

ν … linear packing Density

$\langle R^2 \rangle = Na^2 = N_{BP}a/\nu$

$a/\nu = 2\ nm^2/bp$ (graph)

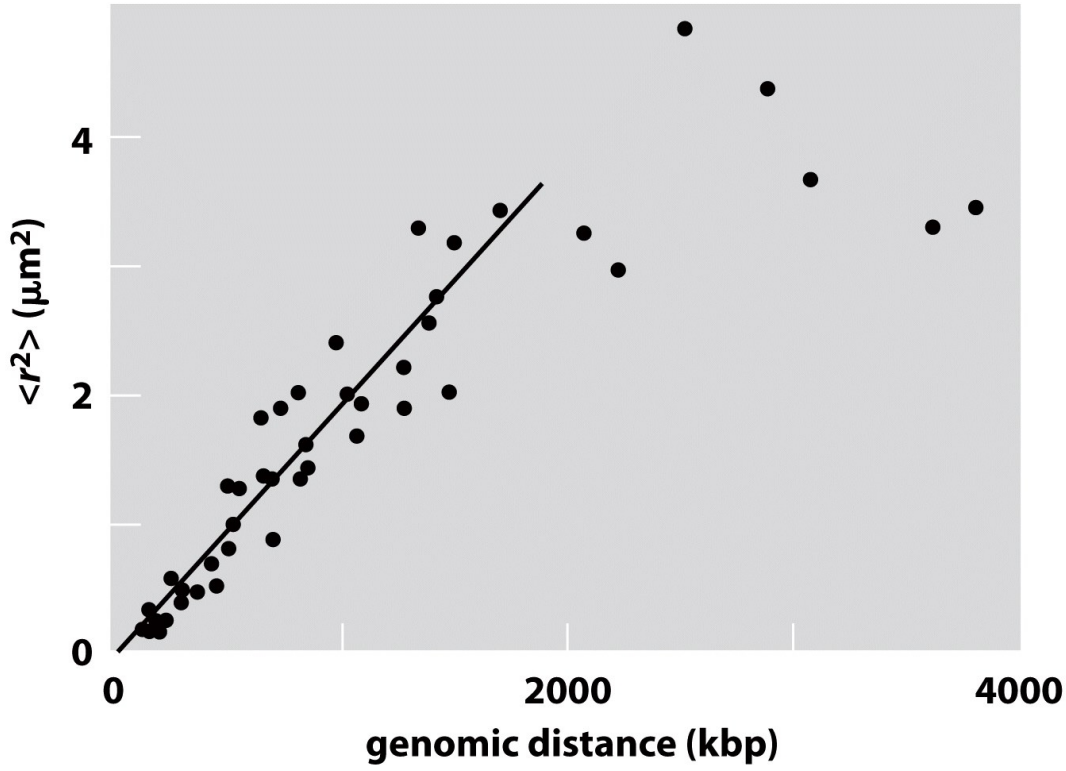


Figure 8.11 Physical Biology of the Cell (© Garland Science 2009)

Measurements with or without tethering using two fluorescently labeled sites results in distributions of distances |r| = r not vectors:

$$P(r) = C\, 4\pi r^2 \exp(-r^2/\sigma^2)$$

for a Gaussian distribution and

$$P(r) = C''\, r/R \{\exp[-(r-R)^2/\sigma'^2] + \exp[-(r+R)^2/\sigma'^2\}$$

$$C'' = C'^{1/3}$$

for a displaced Gaussian distribution.

Can we experimentally distinguish between the Gaussian and displaced Gaussian distribution?

YES, thus, we can detect *in vitro* tethering of chromosomes in cells.

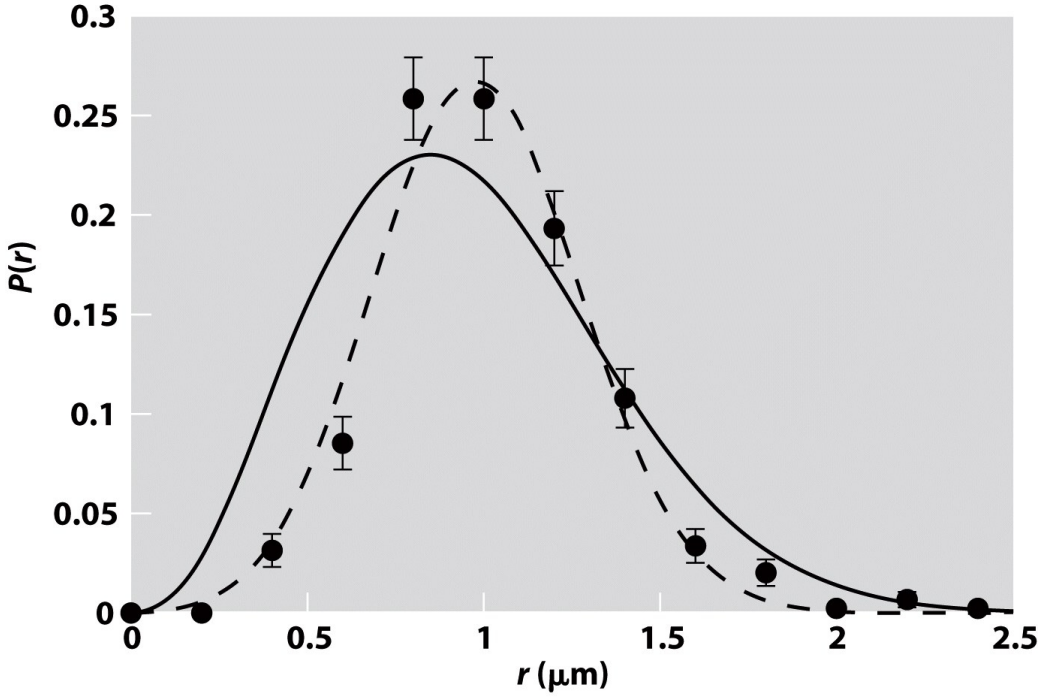
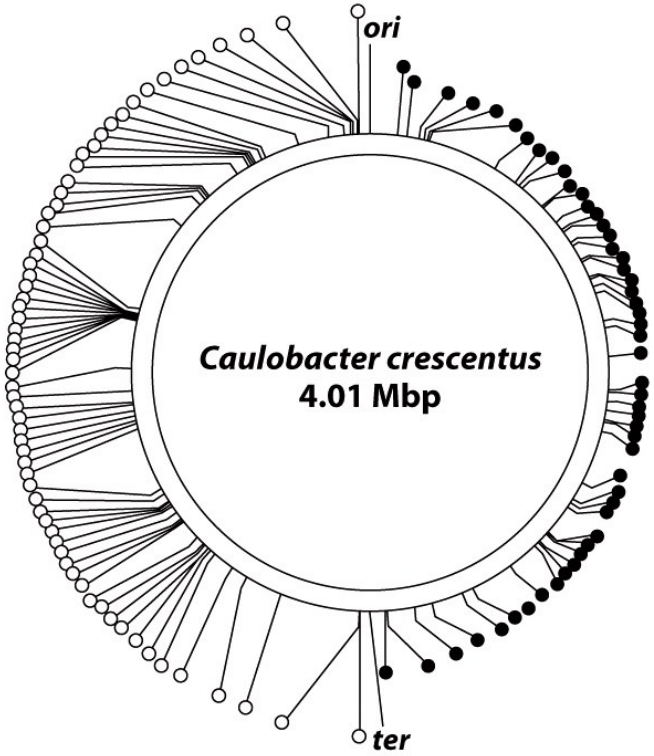


Figure 8.12 Physical Biology of the Cell (© Garland Science 2009)

Chromosome territories in a bacterial cell: Fluorescent tags placed at 112 locations covering the length of the circular chromosome (chromosome partitioned in loops)



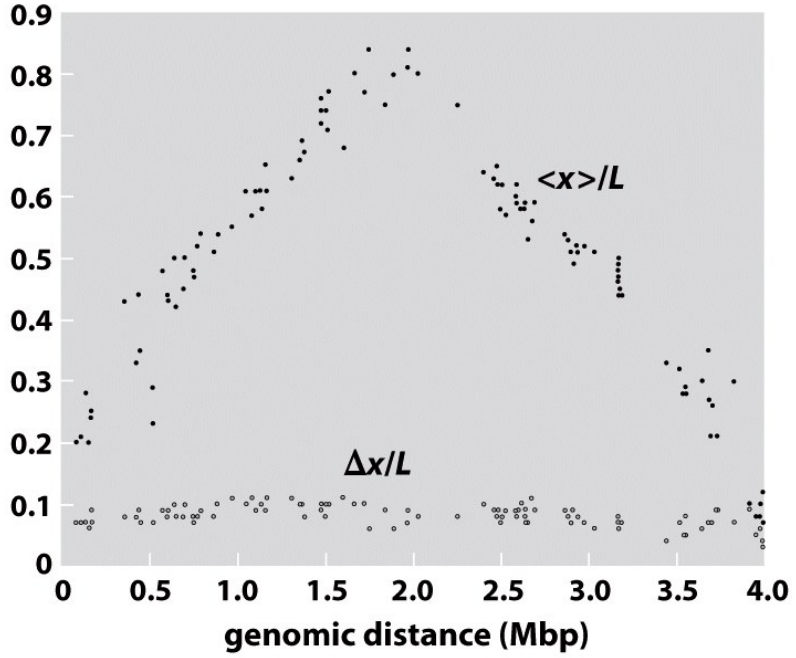
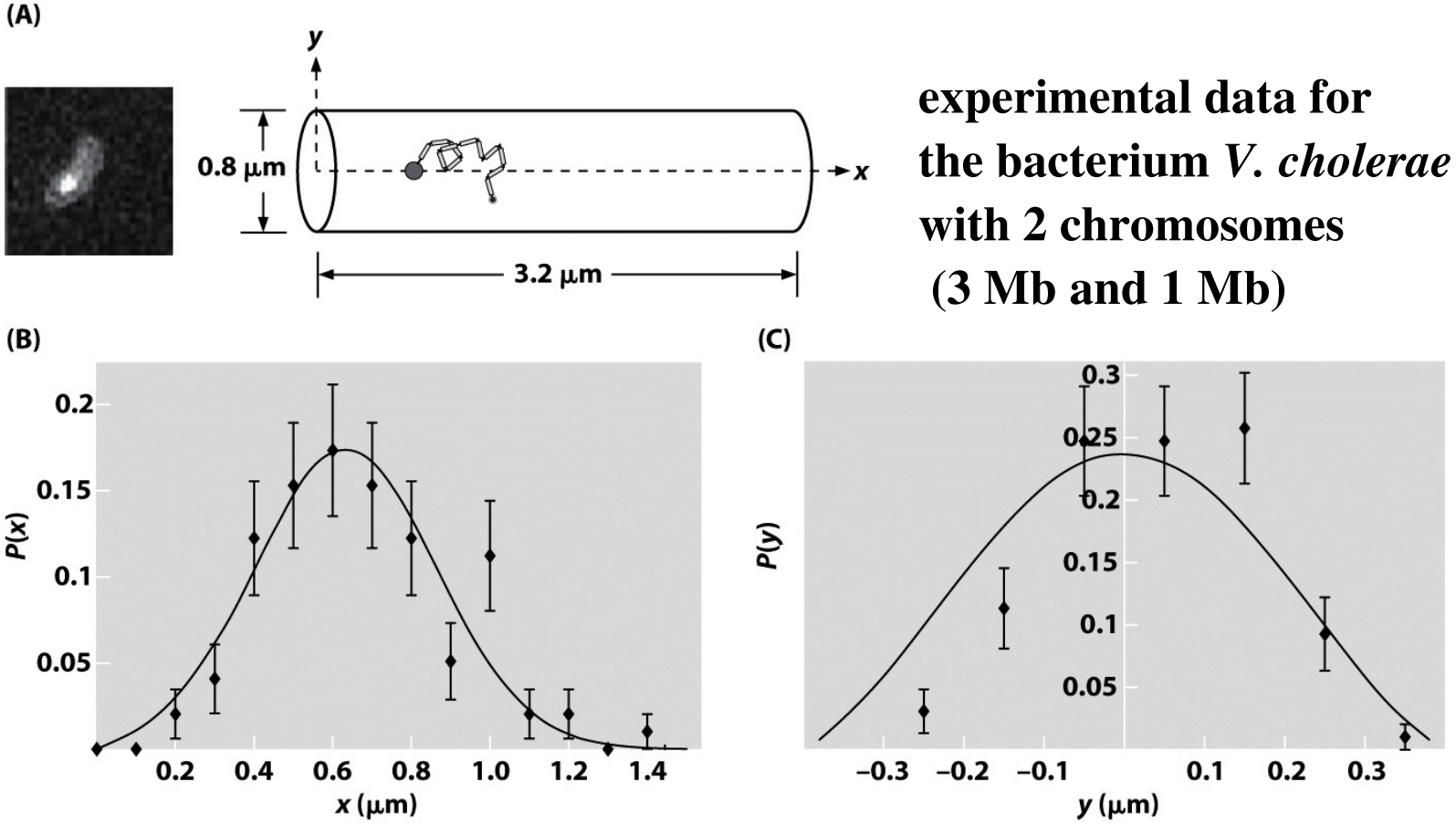


Figure 8.13 Physical Biology of the Cell (© Garland Science 2009)

The result shows a linear relationship between the physical and genomic distances.

Model of polymer confinement and tethering: Gaussian distribution for end-to-end distance (along x) but confined random walk distribution along y



experimental data for the bacterium *V. cholerae* with 2 chromosomes (3 Mb and 1 Mb)

Figure 8.14 Physical Biology of the Cell (© Garland Science 2009)

Random walk model in 1D:

$$P(x) = C \exp[-(x-x_0)^2/\sigma^2]$$

$$C = [1/(2\pi Na^2)]^{1/2} \;\&\; \sigma^2 = Na^2$$

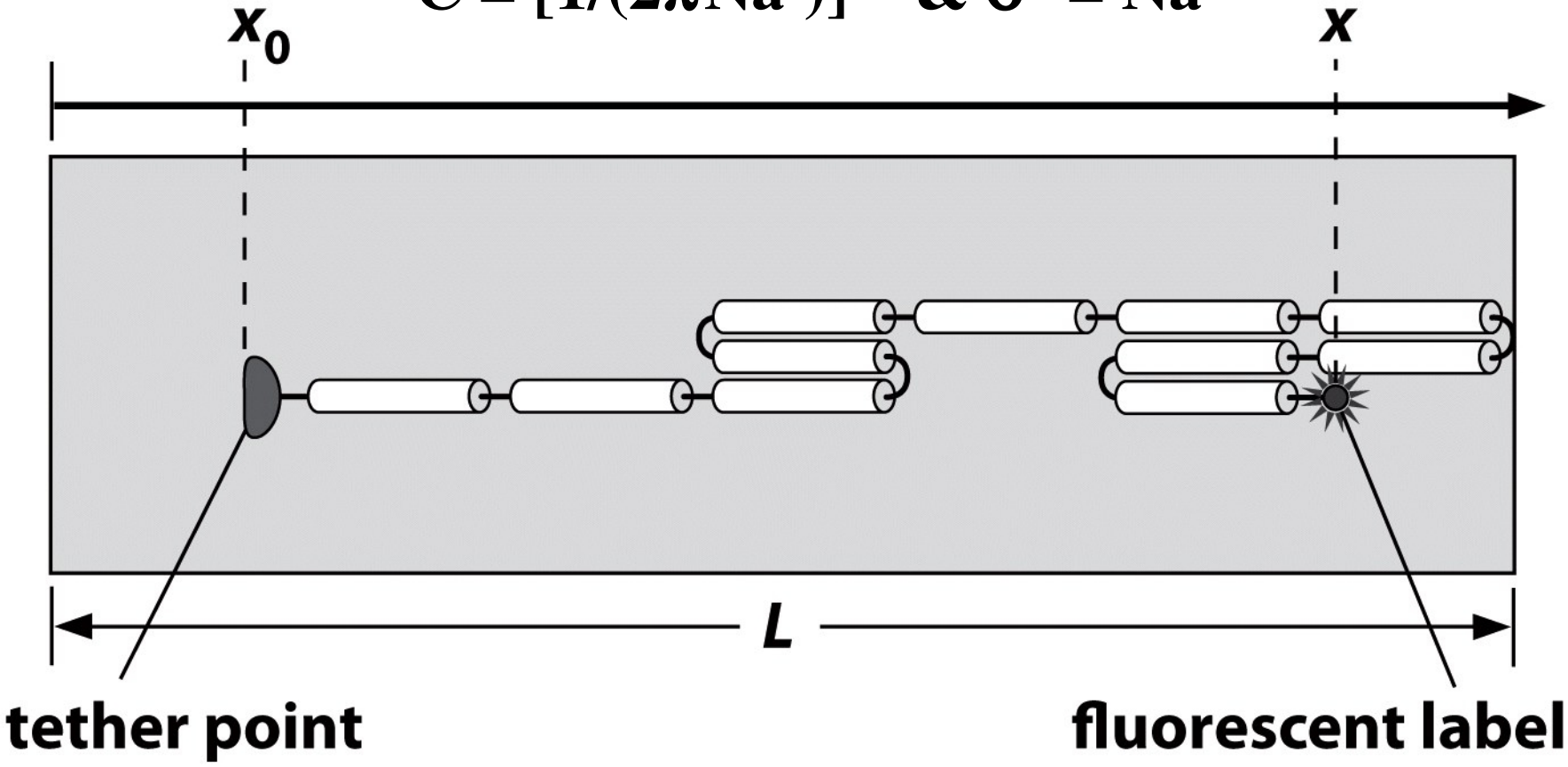


Figure 8.15 Physical Biology of the Cell (© Garland Science 2009)

Random walk problem maps onto the diffusion equation:

Distributions for confined polymers for
(A) different contour lengths of chromosomes (cell size 2 μm)
(B) 1 μm long chromatin fiber confined in cells of different sizes

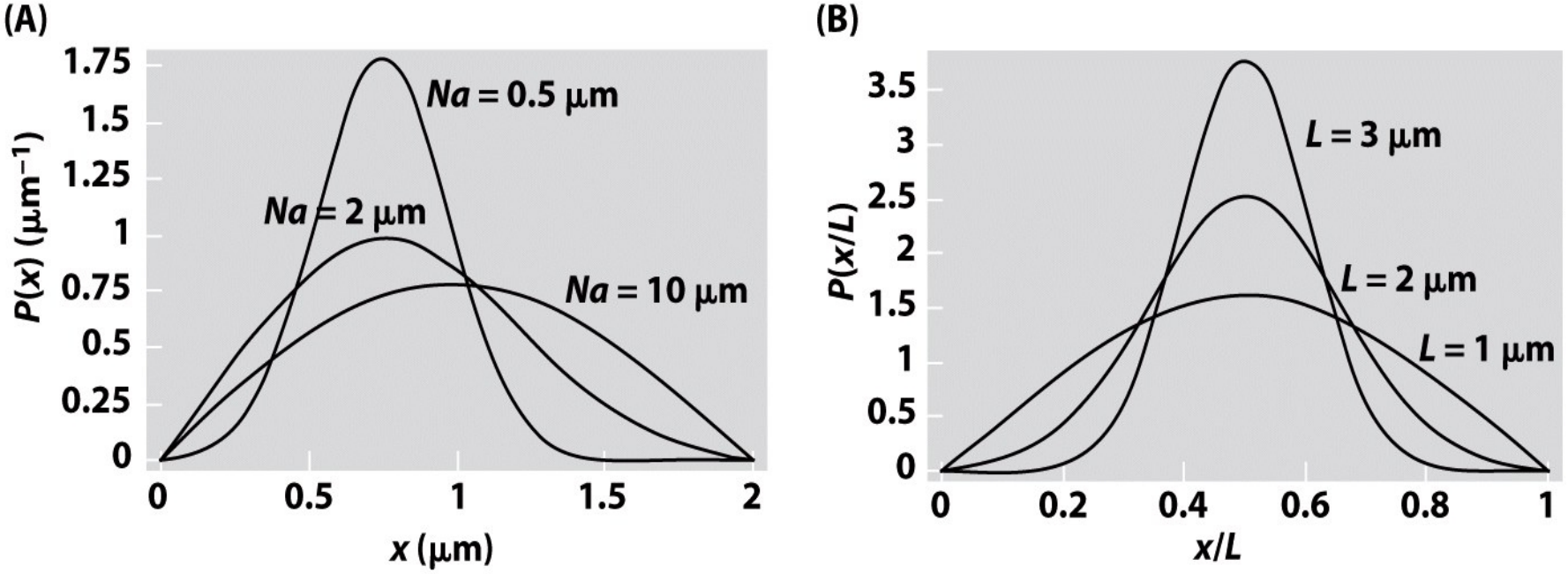


Figure 8.16 Physical Biology of the Cell (© Garland Science 2009)

(A) DNA bubbles, (B) RNA hairpins, © DNA loop, (D) long-range looping of chromosomal DNA

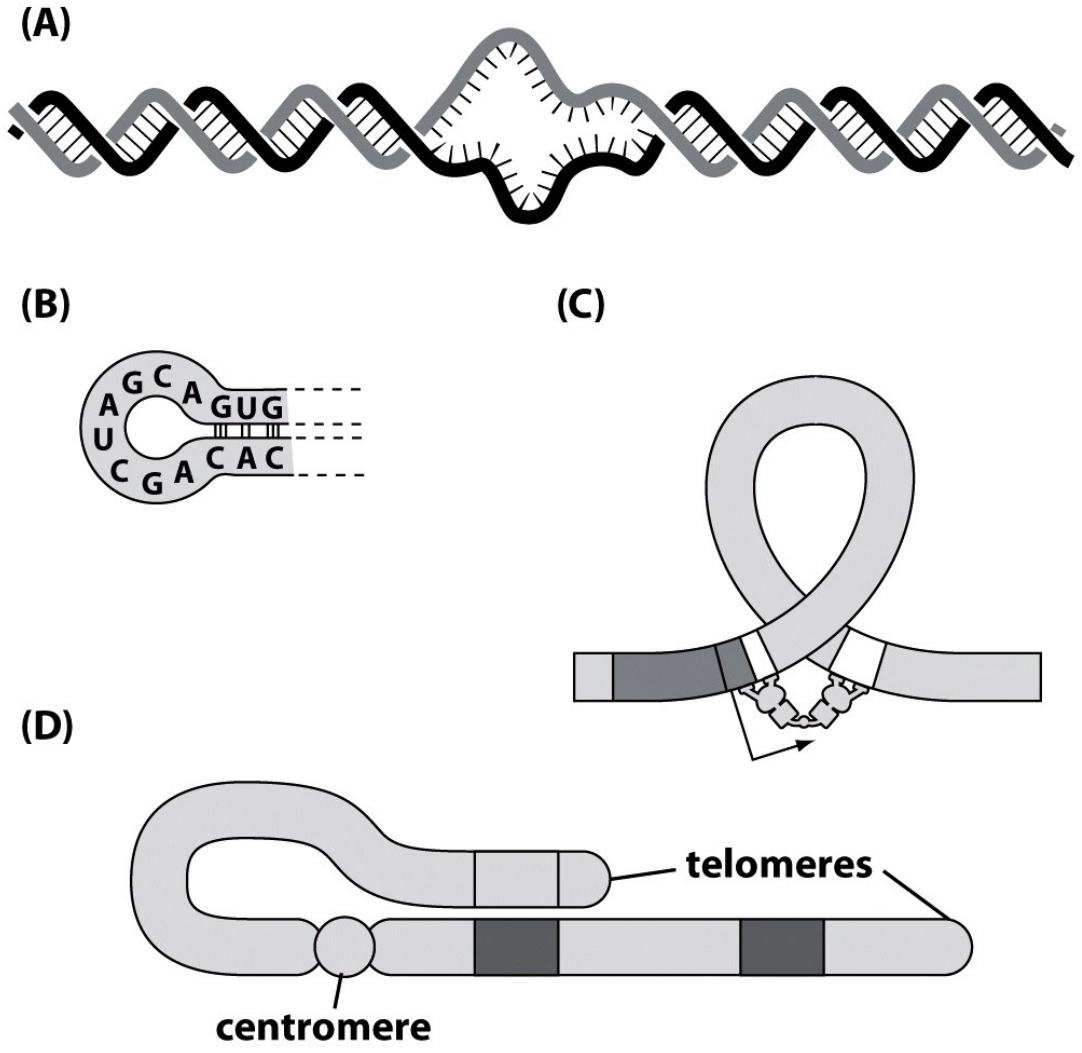


Figure 8.18 Physical Biology of the Cell (© Garland Science 2009)

Model for DNA loop formation by the Lac repressor: Control of protein expression

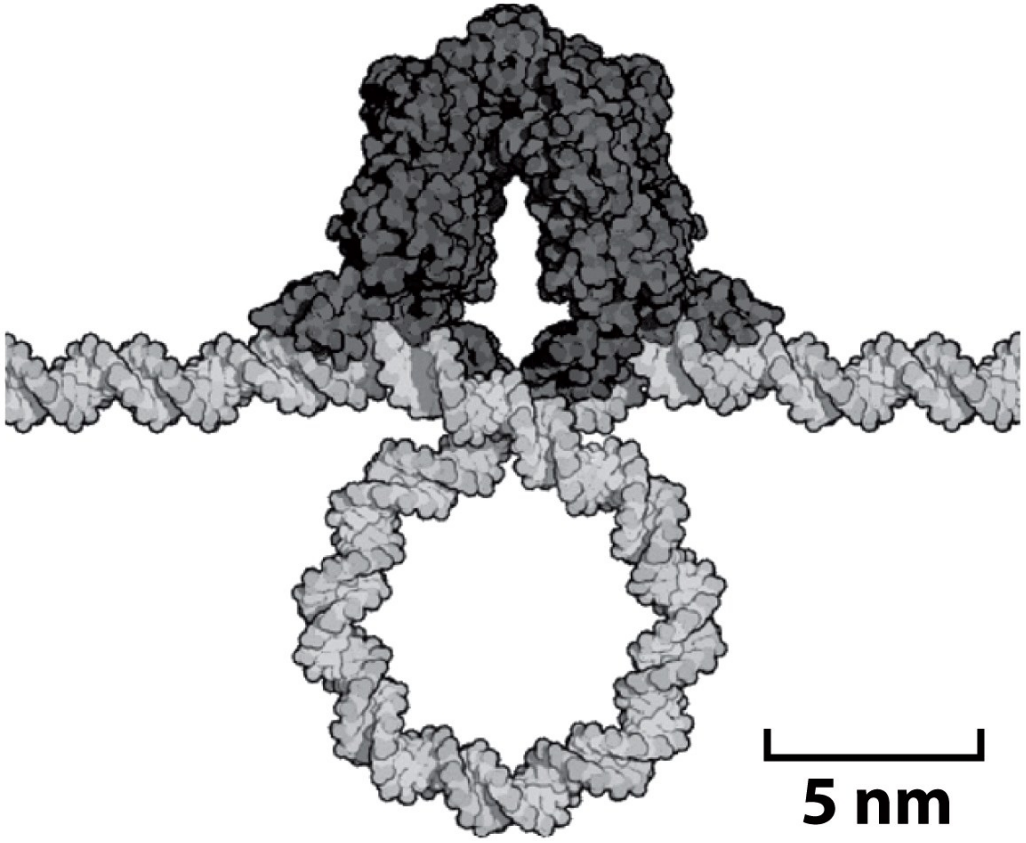


Figure 8.19 Physical Biology of the Cell (© Garland Science 2009)

Entropic cost of loop formation

(1) in 1D random walk model, calculate the fraction of conformations which close on themselves:

$$p_0 = \#\text{ of looped configs / total \# of configs}$$

$$= \{N!/[(N/2)!\,(N/2)!]\}/2^N$$

$$\sim [2/(\pi N)]^{1/2}$$

or use the probability distribution in 1D or 3D

$P(R; N) \sim (2\pi\, Na^2)^{-1/2}$ or $P(R; N) \sim [3/(2\pi\, Na^2)]^{3/2}$

to calculate $p_0 = \int_{-\delta}^{+\delta} P(R; N)\, dR$ and put $\delta = a$

(2) in 3D random walk model,

$$p_0 = \int_0^{+\delta} 4\pi\, R^2\, P(R; N)\, dR = [6/(\pi\, N^3)]^{1/2}$$

PCR = polymerase chain reaction (making many copies of the input DNA fragment)

- **DNA melting at a high T**

- **DNA polymerase at low T makes copies from the primers (20 bp DNA fragments) and nucleotides**

- **the overall concentration of reactant products increases exponentially**

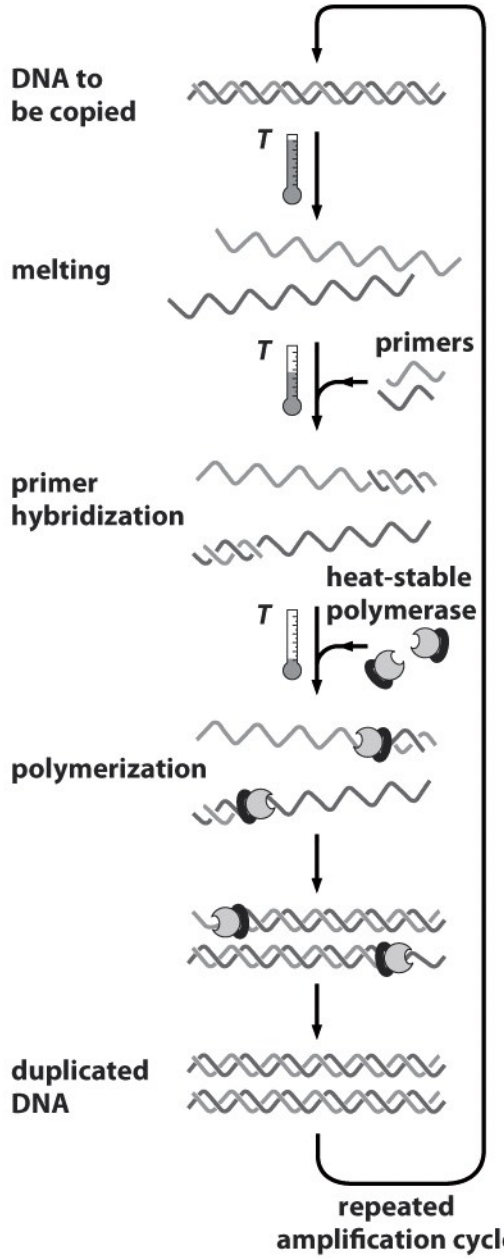


Figure 8.20 Physical Biology of the Cell (© Garland Science 2009)

Formation of DNA bubbles: A competition between entropy maximization and energy minimization

- increases the entropy of the bubble region
- increases the energy due to breaking of favorable hydrogen bonds
- random walk model: a good estimate of the melting T

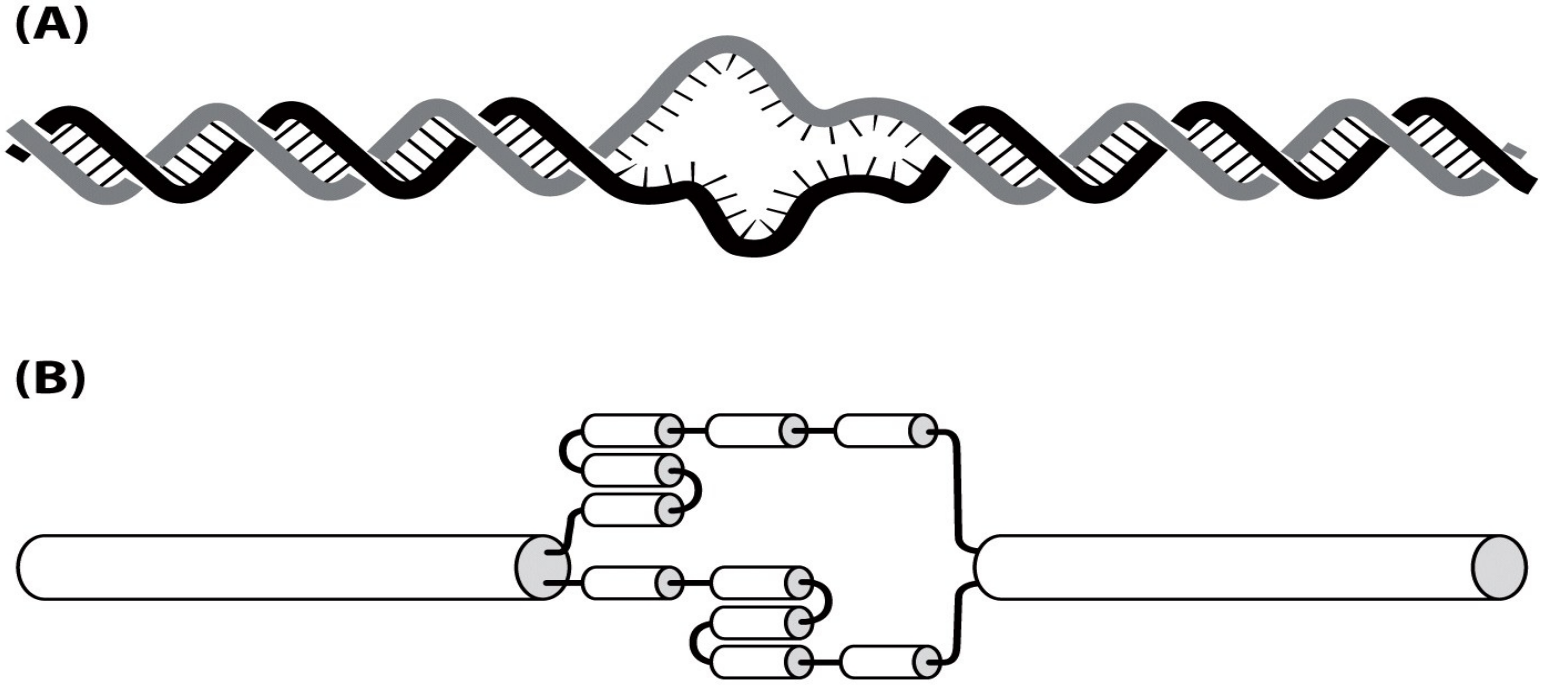


Figure 8.21 Physical Biology of the Cell (© Garland Science 2009)

Calculate that the probability of having a bubble of length n base pairs:

$$p_1(n) = Z^{-1} \exp[-\Delta G_1(n)/k_B T]$$

$\Delta G_1(n)$ is the free energy for formation of a bubble of length n:

$$\Delta G_1(n) = E_{IN} + nE_{EL} - k_B T \ln[\Omega_0(n)\,(N-n+1)]$$

E_{IN} … energy of bubble initiation

E_{EL} … energy of bubble elongation by one base pair

$(N-n+1)$ … number of choices for the bubble location

$\Omega_0(n)$ … the number of ways to make a bubble of two strands each n base pairs long (1D)

$$\Omega_0(n) = 2^{2n}\, p_0(2n) \sim 2^{2n}/(\pi N)^{1/2} \text{ for } n \gg 1$$

$$\Delta G_1(n)/k_BT = n(\varepsilon_{EL} - 2\ln 2) + \tfrac{1}{2}\ln n - \ln(N-n+1)$$

where $\varepsilon_{EL} = E_{EL}/k_BT$

Minimize $\Delta G_1(n)/k_BT$ with respect to n:

$$\varepsilon_{EL} - 2\ln 2 + 1/(2n) + 1/(N-n+1) = 0$$

Two possible situations:

(A) $\varepsilon_{EL} - 2\ln 2 > 0$: no real solution for n, bubbles small

(B) $\varepsilon_{EL} - 2\ln 2 < 0$: two solutions for n (small n local maximum and n~N local minimum)

(A) low temperatures; (B) high temperatures

Interplay between the energy cost of bubble elongation and thermal energy: $\varepsilon_{EL} = E_{EL}/k_B T$

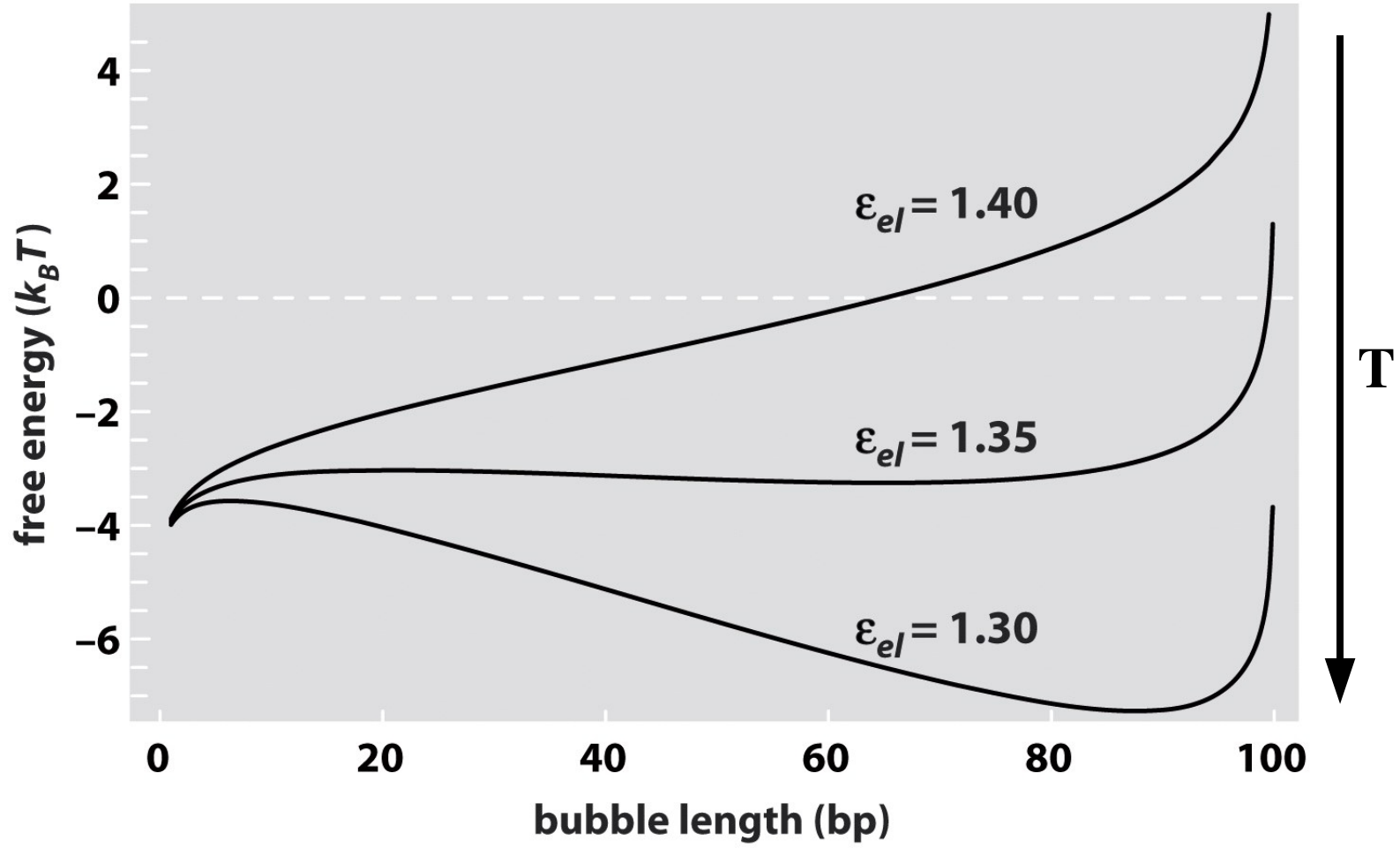


Figure 8.22 Physical Biology of the Cell (© Garland Science 2009)

Single Molecule Experimental Techniques

A) AFM

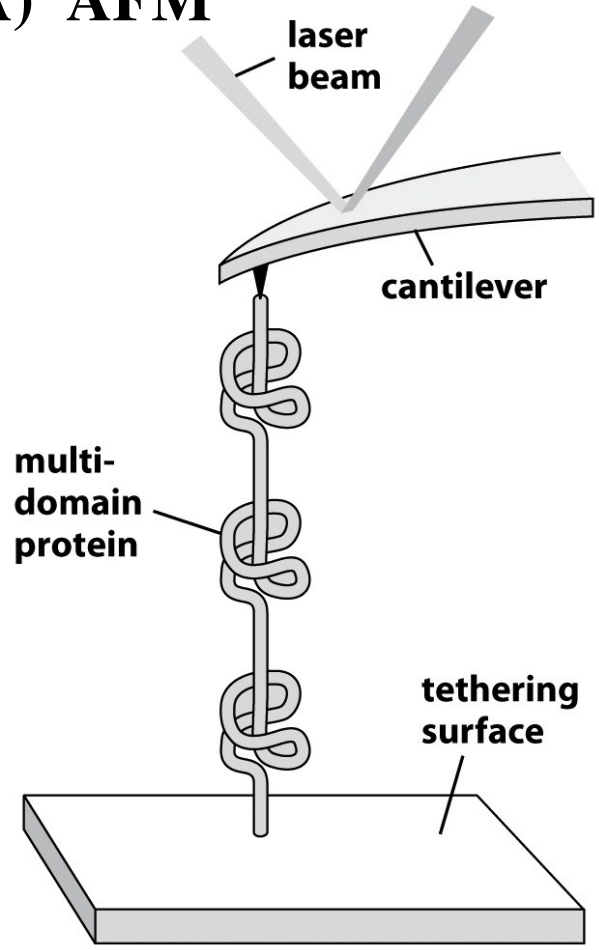


Figure 8.23a Physical Biology of the Cell (© Garland Science 2009)

B) optical tweezers

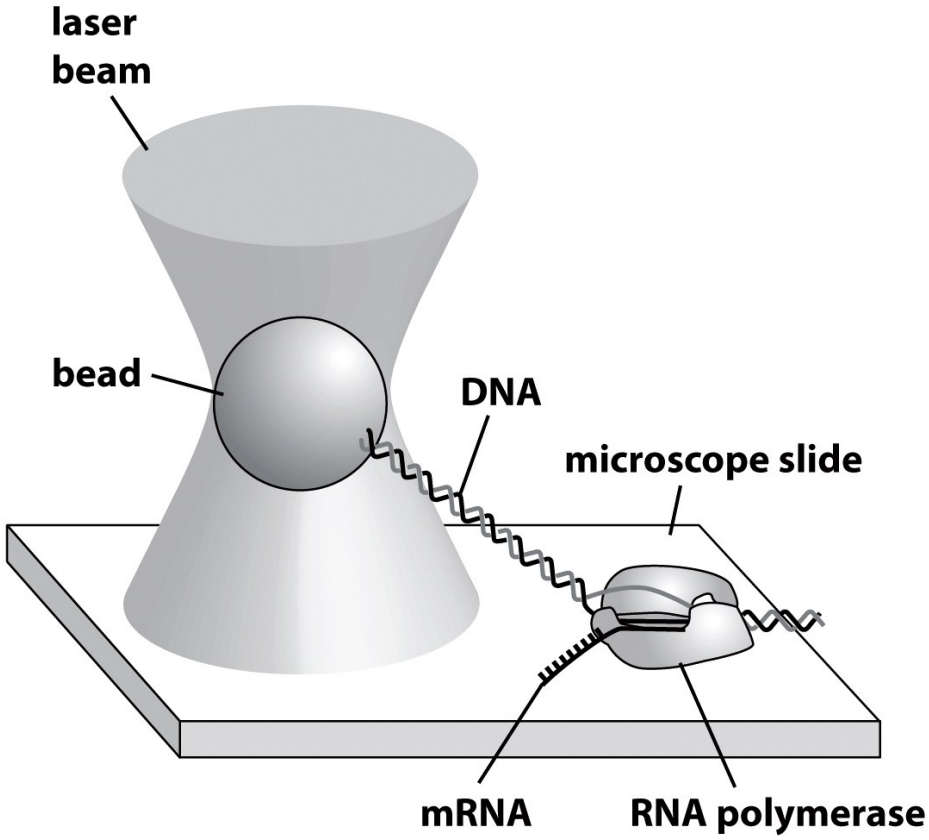


Figure 8.23b Physical Biology of the Cell (© Garland Science 2009)

C) magnetic tweezers

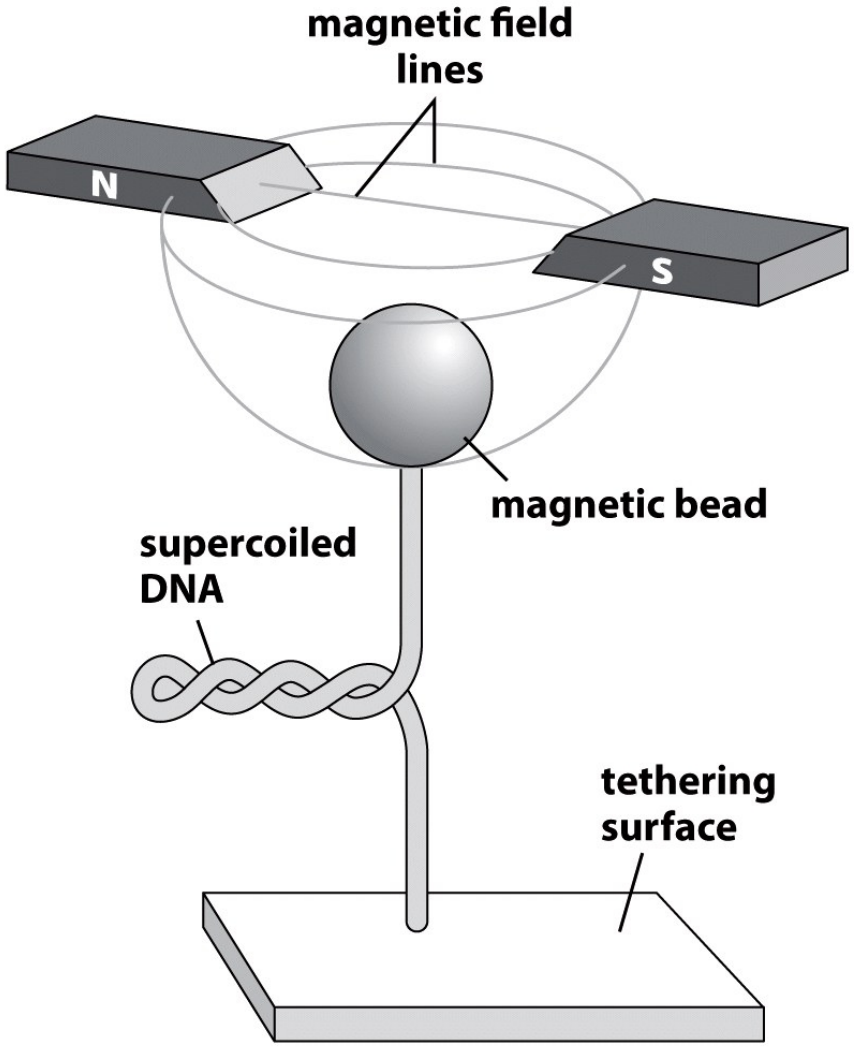


Figure 8.23c Physical Biology of the Cell (© Garland Science 2009)

D) pipette based force apparatus (ligand-receptor adhesion forces)

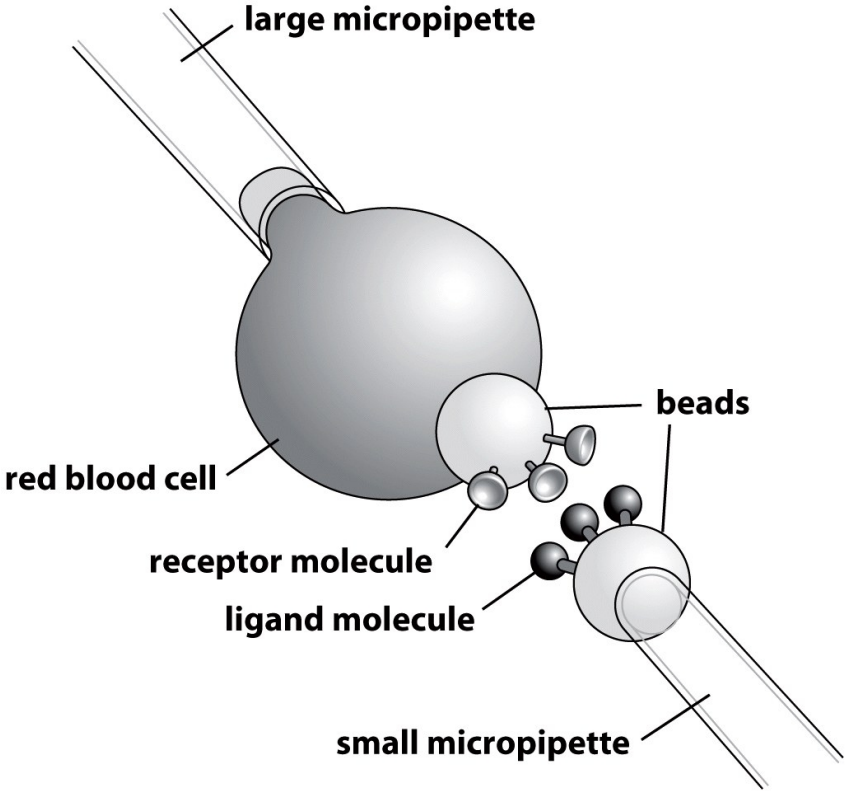
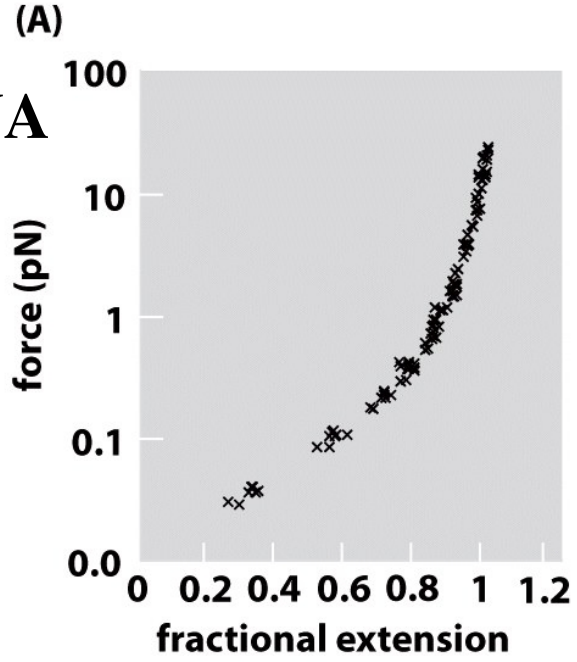
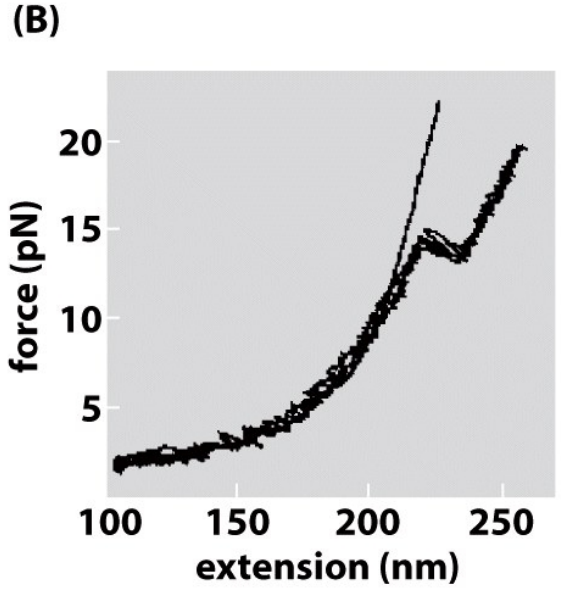


Figure 8.23d Physical Biology of the Cell (© Garland Science 2009)

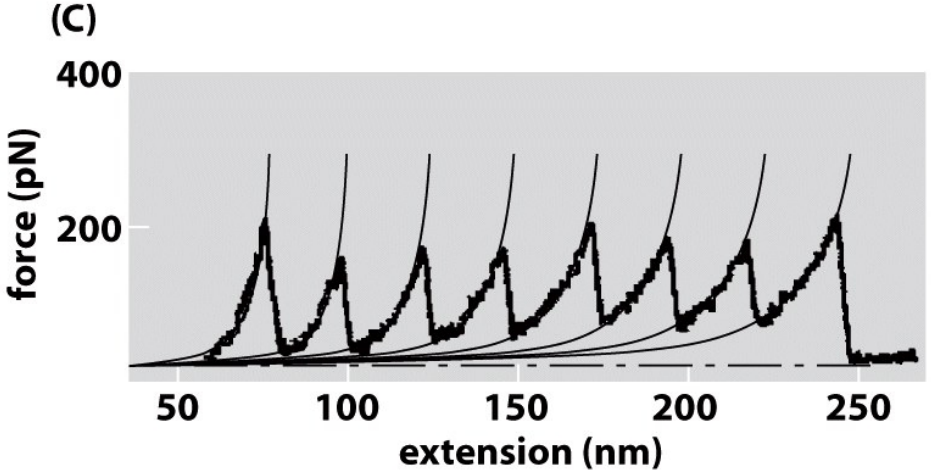
Force-extension curves: fingerprints of specific macromolecules

single DNA





RNA



titin (protein with multiple domains)

Figure 8.24 Physical Biology of the Cell (© Garland Science 2009)

Random Walk Models for Force-Extension Curves

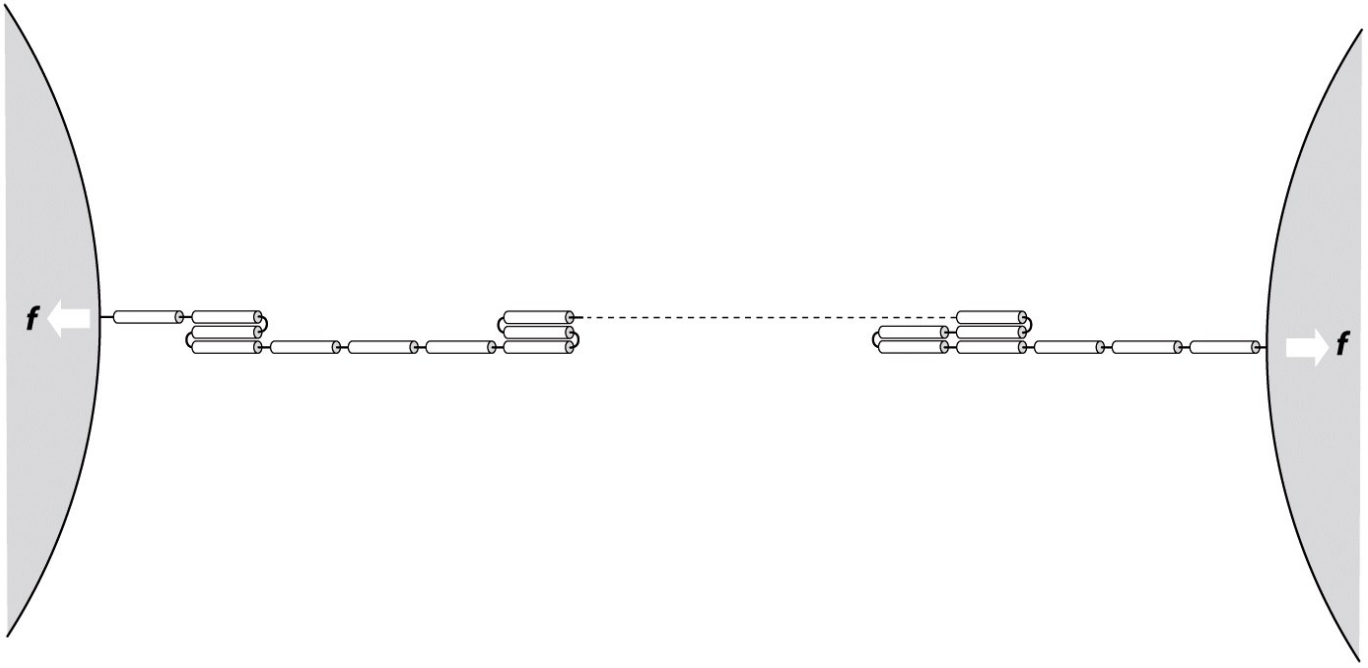


Figure 8.25 Physical Biology of the Cell (© Garland Science 2009)

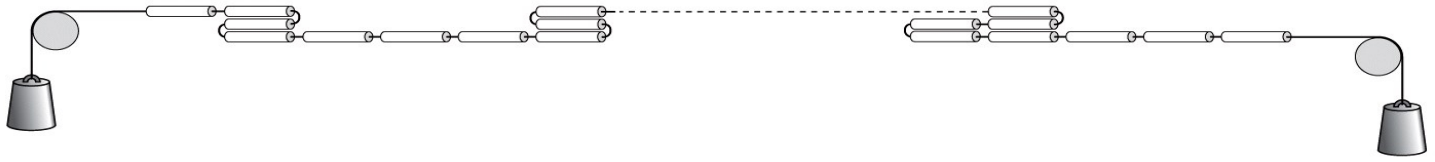


Figure 8.26 Physical Biology of the Cell (© Garland Science 2009)

The free energy of stretching with a force f can be expressed as:

$$G(L) = - f L - k_B T \ln W(L; L_{TOT})$$

L … end-to-end distance of the macromolecule

L_{TOT} … total length of the macromolecule

$W(L;L_{TOT})$ … the number of microstates (realizations, permutations) corresponding to L

f … pulling force

In a 1D random walk model:

$$L = (n_R - n_L)\, a \quad \& \quad L_{TOT} = (n_R + n_L)\, a = N a$$

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

The free energy can be expressed as:

$$G(n_R) = -2fn_R a - k_B T\,[n_R \ln(n_R) + (N-n_R)\ln(N-n_R)]$$

$$\partial G(n_R)/\partial n_R = 0 \Rightarrow \; n_R/n_L = \exp[2fa/(k_B T)]$$

The 1D random walk model predicts the extension z:

$$z = \langle L\rangle/L_{TOT} = (n_R - n_L)/(n_R + n_L) = \tanh(fa/k_B T)$$

For $fa \ll k_B T$ we obtain a linear relationship

$$\langle L\rangle = L_{TOT}\, fa/(k_B T)$$

Hooke's law: $f = kx$, thus the stiffness constant:

$$k = k_B T/(aL_{TOT}) \ldots \text{entropic origin}$$

Force-Extension Curves Predicted by the Freely Joined Chain Model: Comparison of 1D, 2D, 3D, and 3D off-lattice models

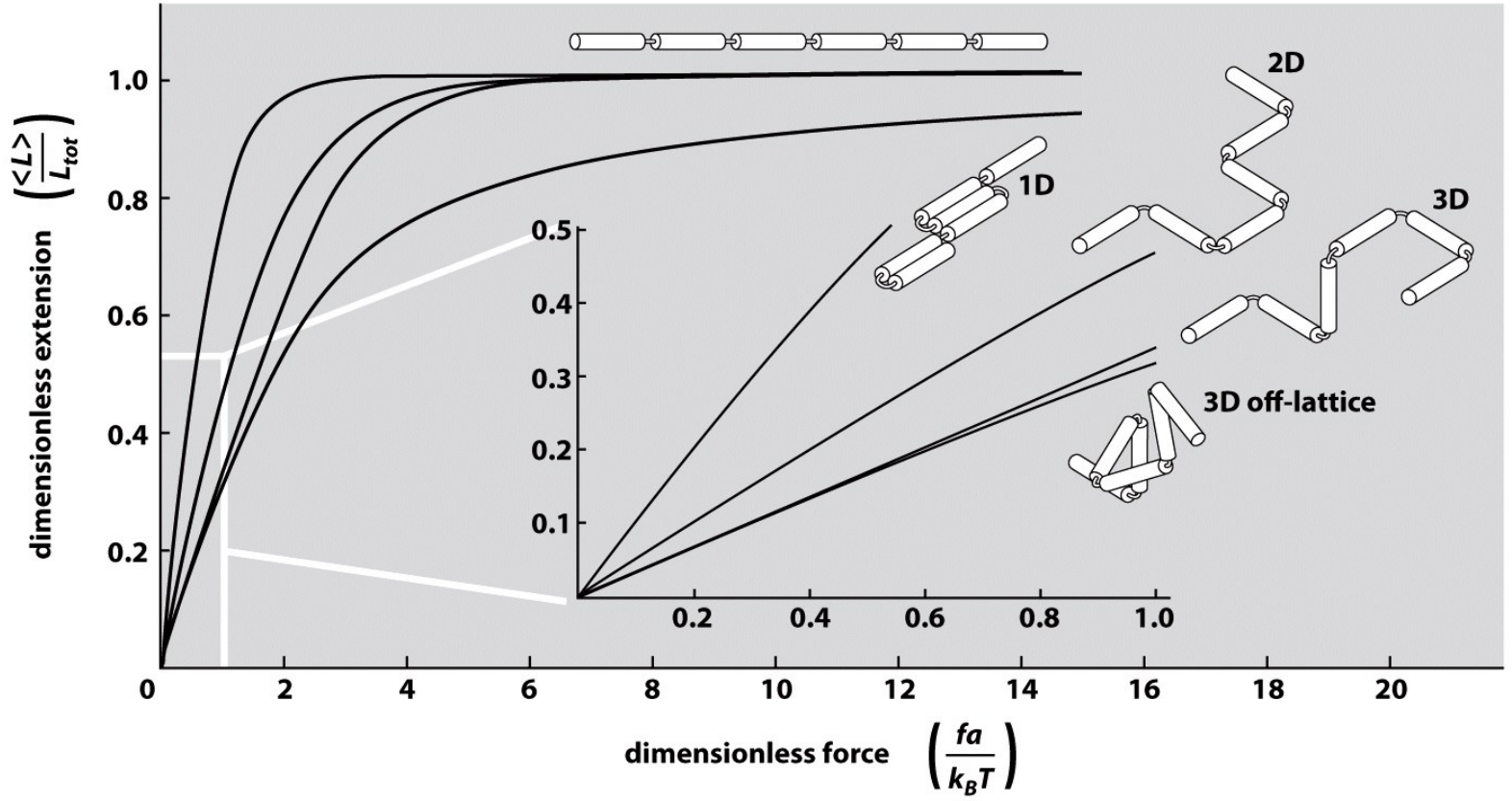


Figure 8.27 Physical Biology of the Cell (© Garland Science 2009)

Proteins as Random Walks

- globular proteins more compact than the RW model predictions

- compact self-avoiding random walk model (filling up the lattice)

- globular proteins scaling:

$$R_G \sim m^{1/3} \sim N^{1/3}$$

(N … # of amino acids)

- unfolded proteins (random coils) scale as: $R_G \sim N^{1/2}$ (RW) or $R_G \sim N^{3/5}$ (self-avoiding RW)

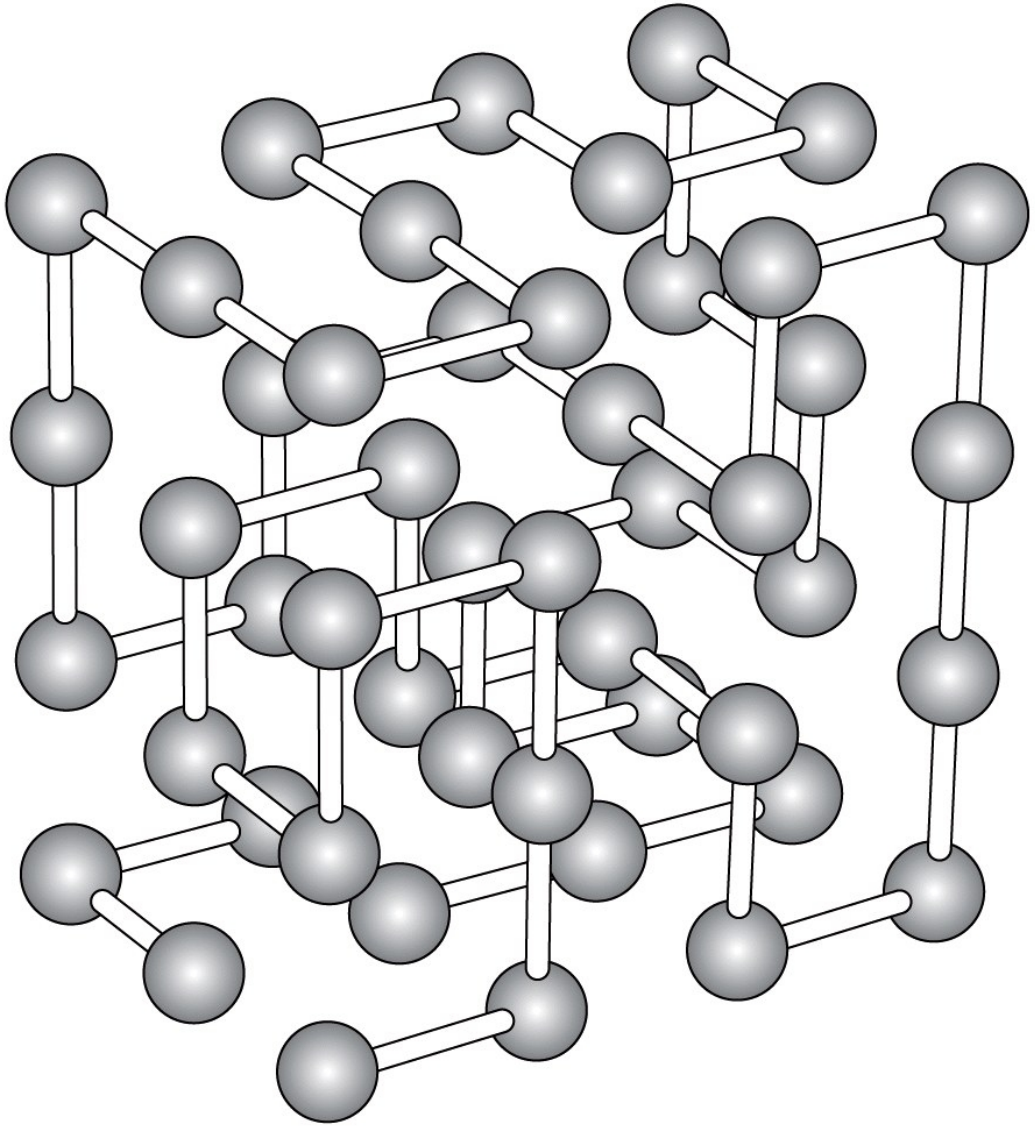


Figure 8.28 Physical Biology of the Cell (© Garland Science 2009)

Scaling of protein size (radius) with the number of amino acids

Slope = 1/3: space filling packing of globular proteins

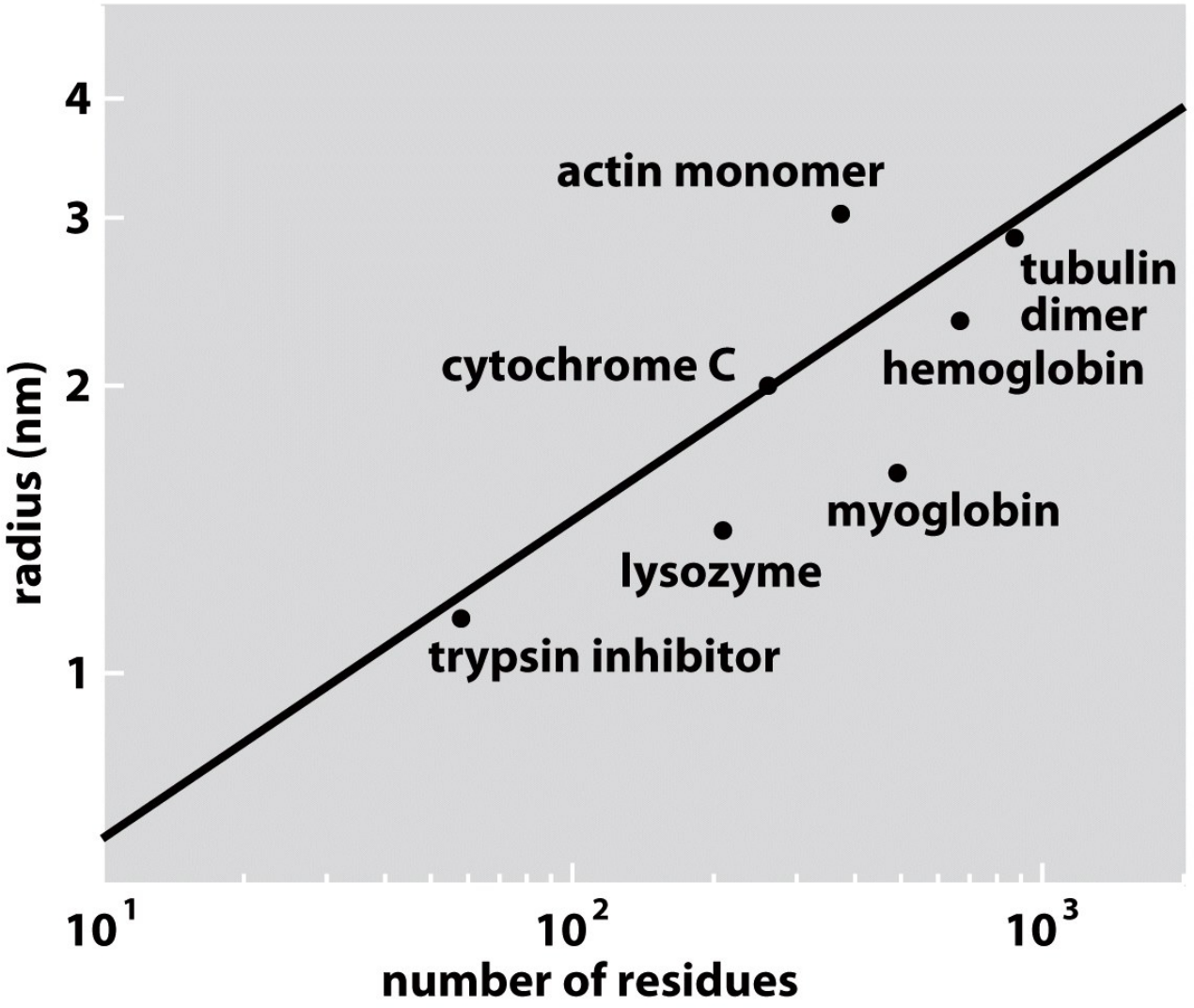


Figure 8.29 Physical Biology of the Cell (© Garland Science 2009)

Levinthal's paradox:

- too many ways to fold a protein of ~100 amino acids yet proteins in nature find their native conformations fast (μs to ms)

- in a compact RW model on a 3D lattice 6^{100} different conformations possible: ~ 6.5×10^{77}

- to explore all conformations (count 10^{-15} s for each) it would take ~ 2×10^{55} years (10^{45} times the age of the universe)

Protein Modeling Beyond the Random Walk-Type Models: HP Model = Each Residue is Either H or P

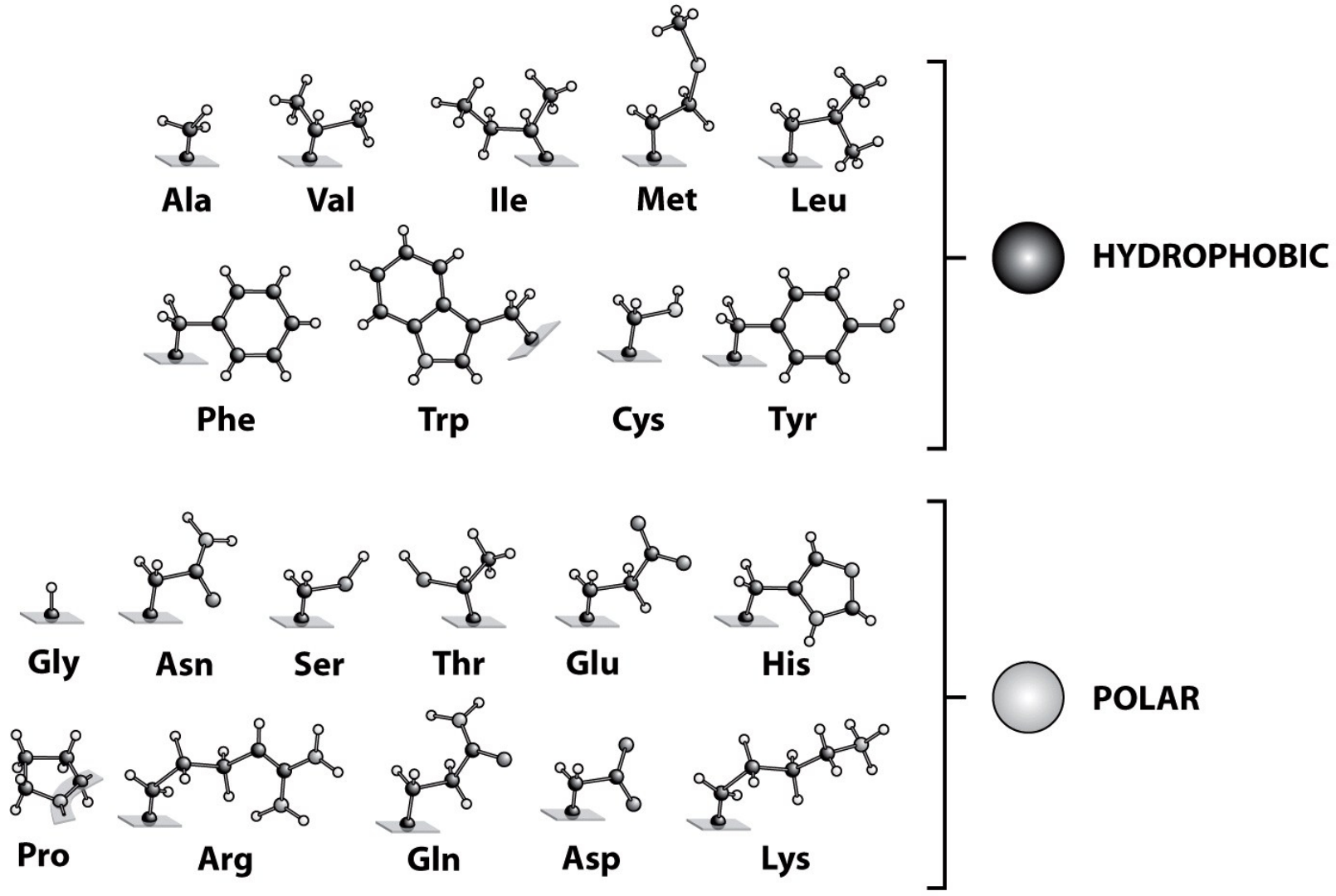


Figure 8.30 Physical Biology of the Cell (© Garland Science 2009)

Classification of the 20 naturally occurring amino acids into 7 classes

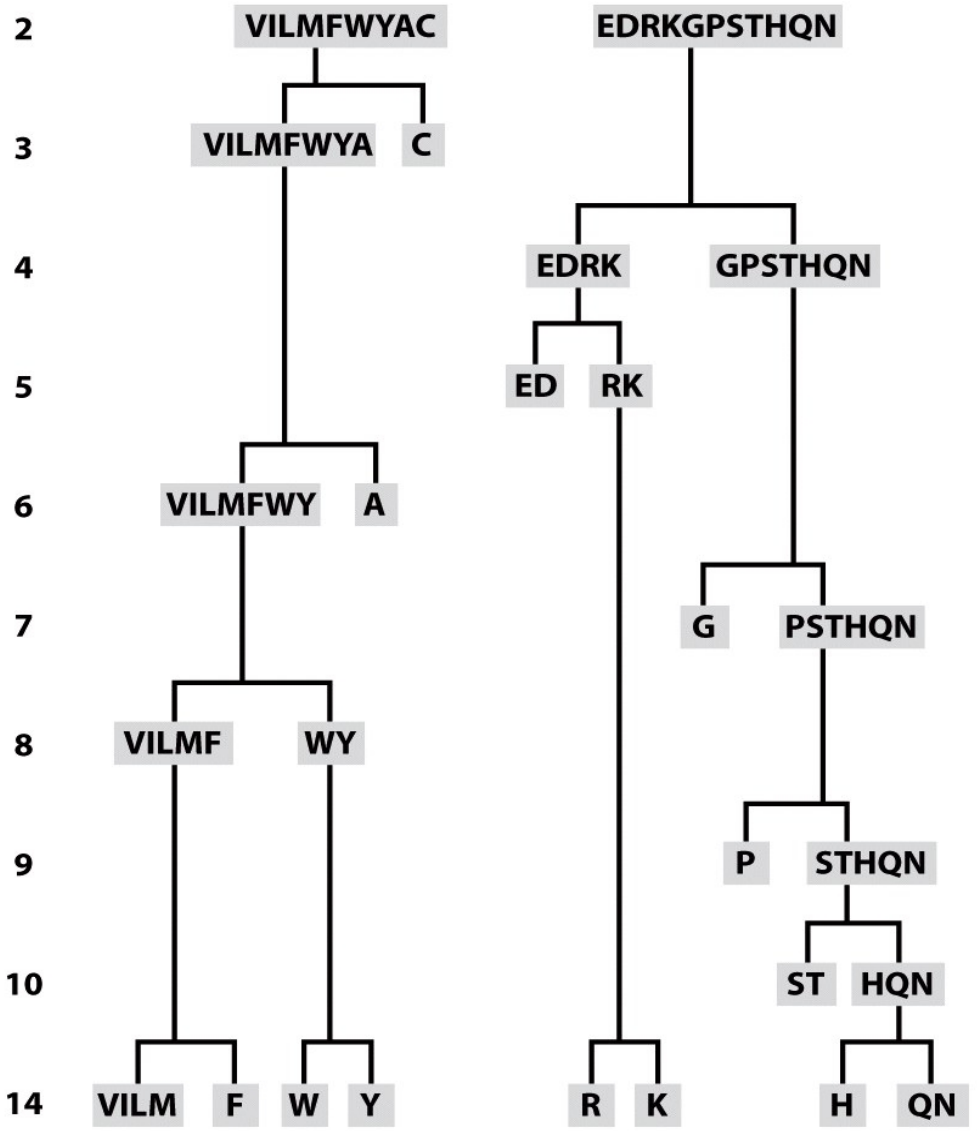


Figure 8.31 Physical Biology of the Cell (© Garland Science 2009)

HP Protein Models

3 x 3 x 3 lattice:

- **103,346 compact structures**
- **2^{27} = 134,217,728 sequences**

2 x 3 lattice (right):

- **2^{6} = 64 sequences**
- **3 possible compact structures (non-convertible by rotations & translations)**

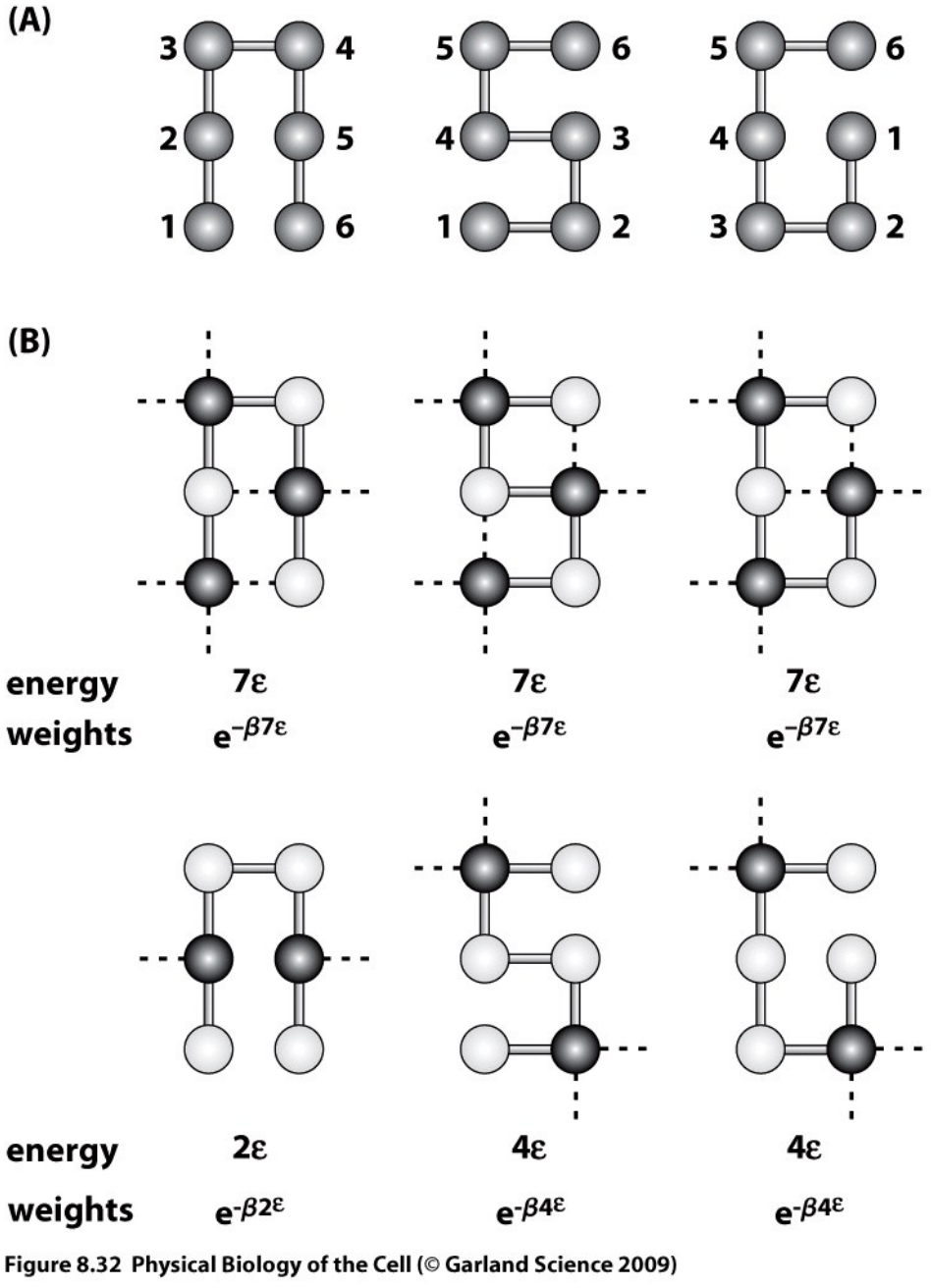


Figure 8.32 Physical Biology of the Cell (© Garland Science 2009)

Energy: assign ε to every unfavorable contact (H with P or solvent); depends on the protein sequence

Probability of the folded state for PHPPHP sequence

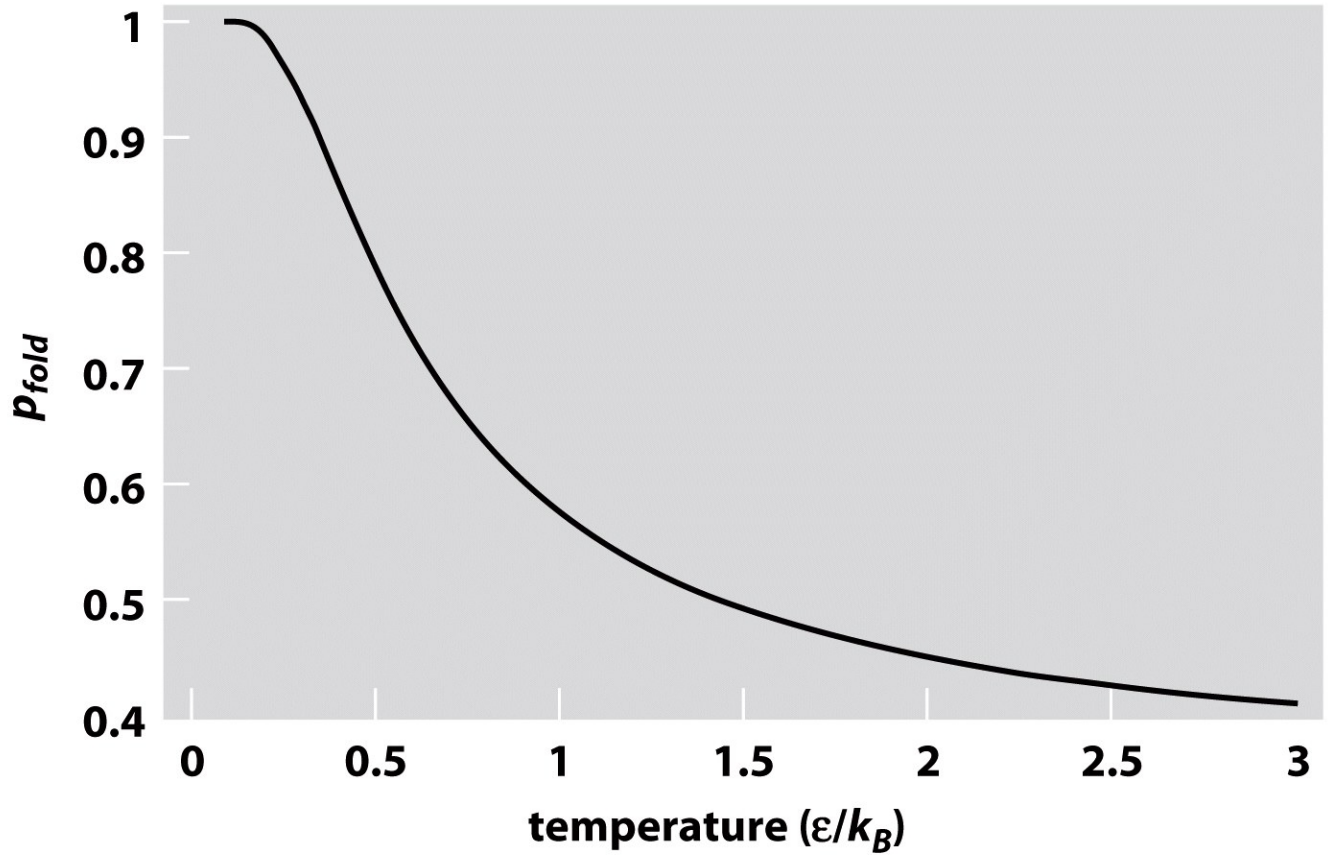


Figure 8.33 Physical Biology of the Cell (© Garland Science 2009)

Not all sequences have a native state (are protein-like)!

Find protein-like sequences (right) with unique compact structure

The structures with a high number of sequences are highly *designable*.

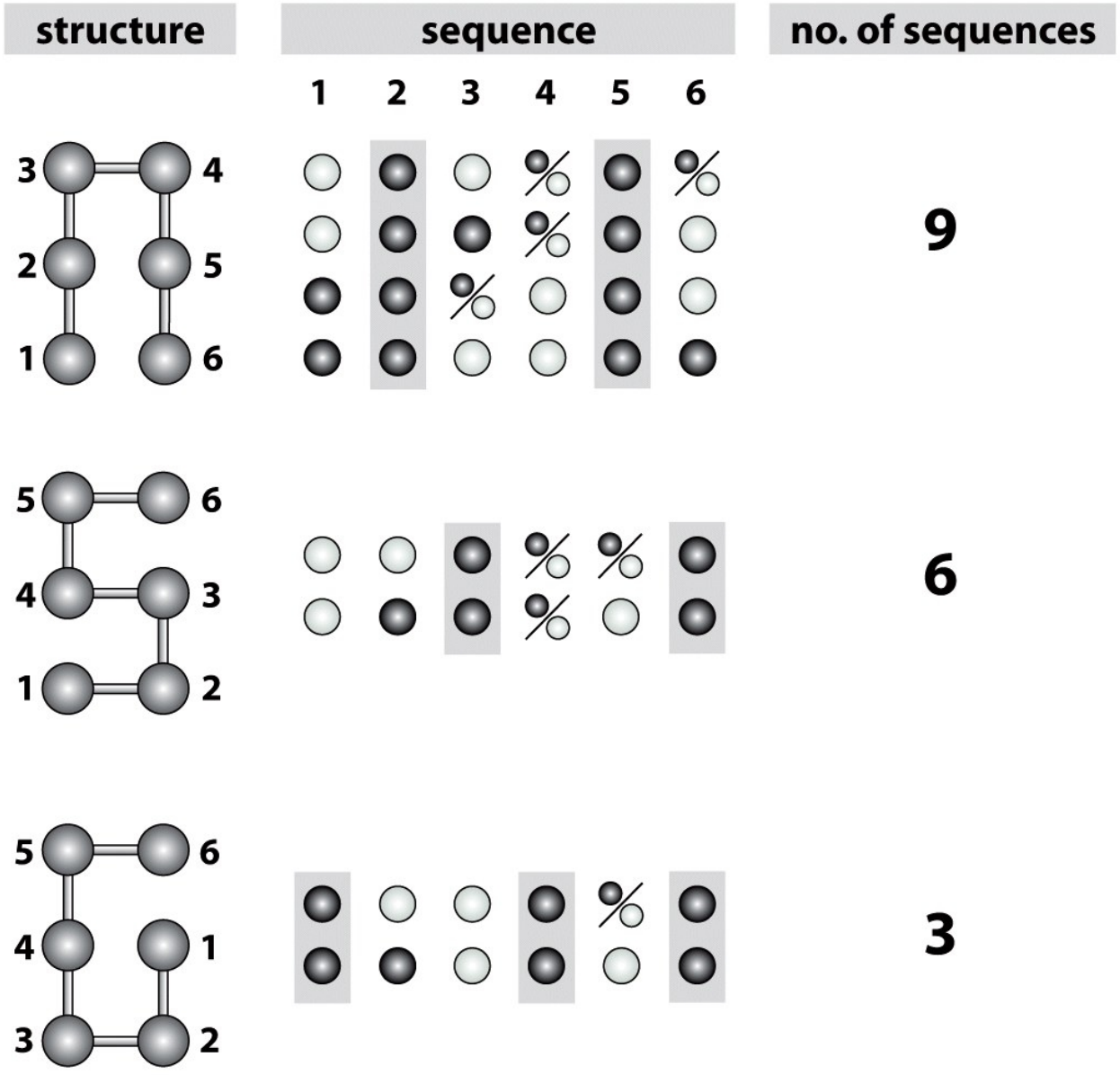


Figure 8.34 Physical Biology of the Cell (© Garland Science 2009)

Four helix bundle protein: HPPHHPPHPPHHPP …
was designed to form a-helices (H every 3-4 residues)

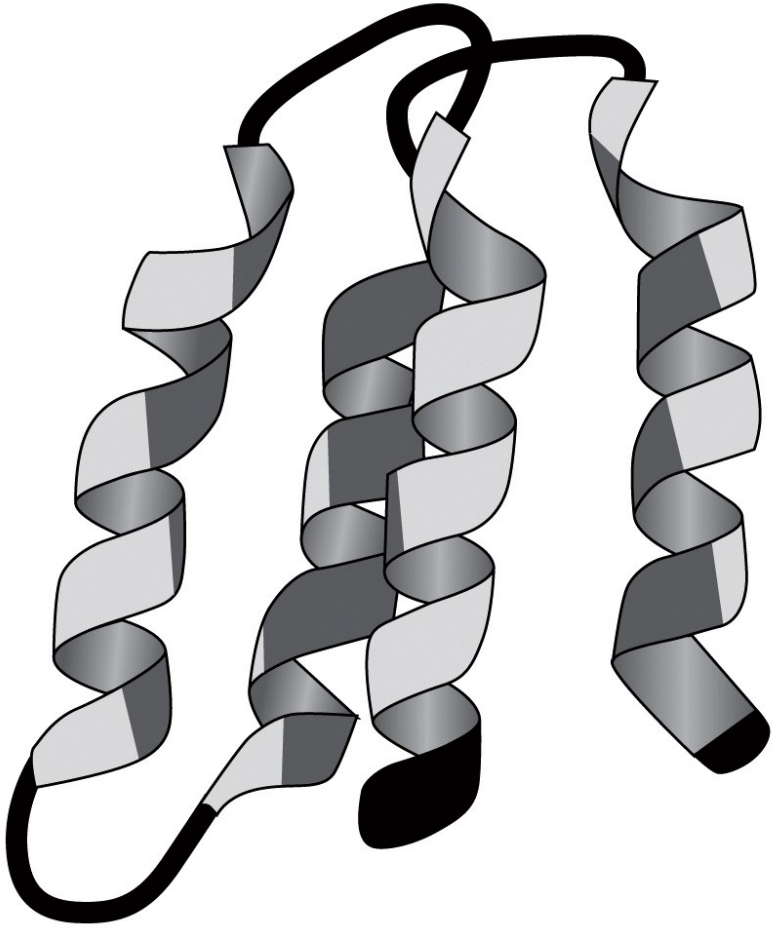


Figure 8.35 Physical Biology of the Cell (© Garland Science 2009)