Zhou Pei-Yuan Centre for Applied Mathematics, Tsinghua University
November 2013

F. Piazza
Center for Molecular Biophysics and University of Orléans, France

Selected topic in Physical Biology

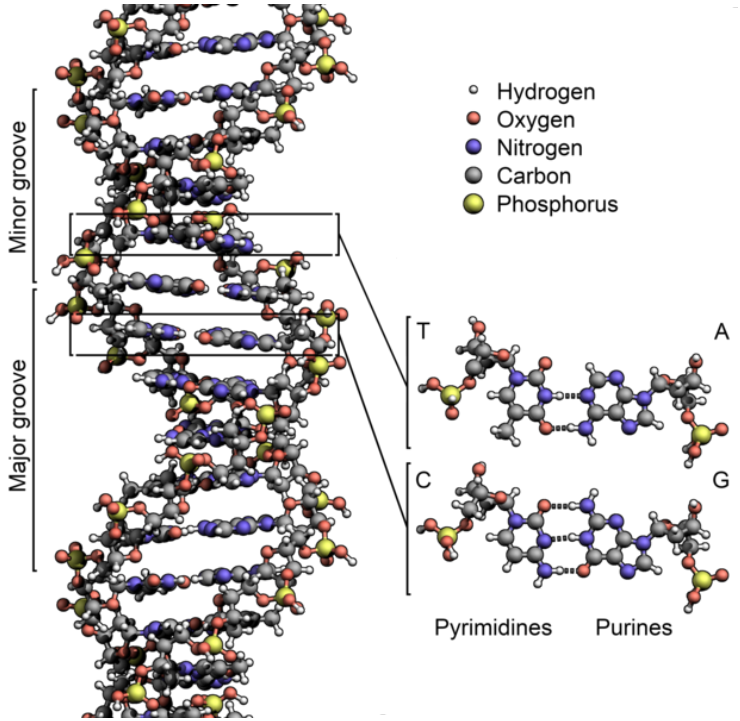
Lecture 2

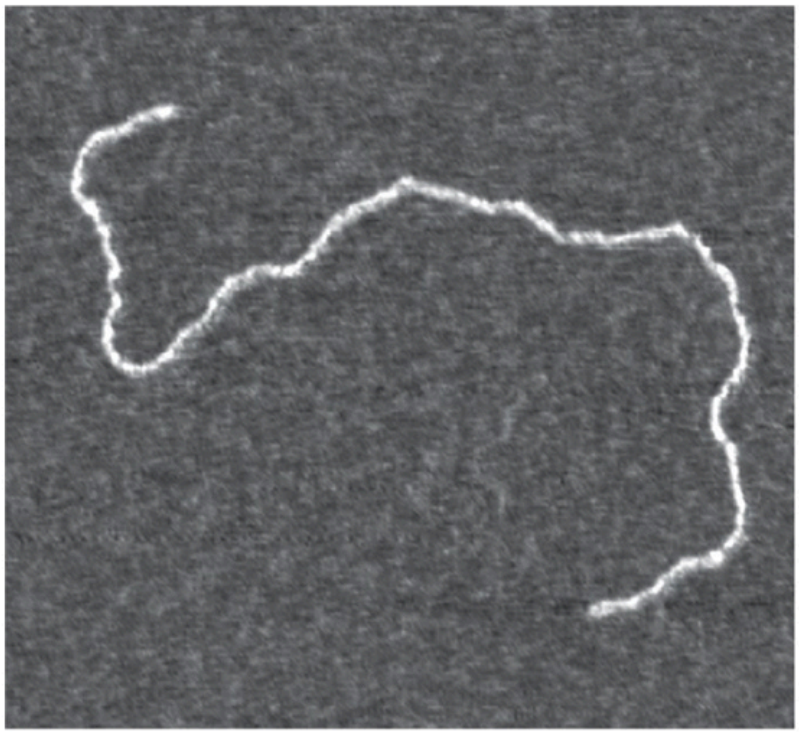
The structure of biological macromolecules: Introduction to statistical models of polymers and applications to DNA stretching

Deterministic versus statistical description of structure

Coordinate (PDB) files reflect a **deterministic** description of macromolecular structure. In a sense it is a deceptively static picture that can only thought of as a starting point

Notion of universality comes immediately when thinking to polymeric systems (e.g. DNA). Think of a virus DNA ejected into the cell, relevant questions are (i) how much space it takes (ii) where in the cell it does so





The random walk model

Deterministic: vector $\boldsymbol{r}(s)$ of position at a distance s along the contour
Statistical: Rigid segments of length a connected by flexible links

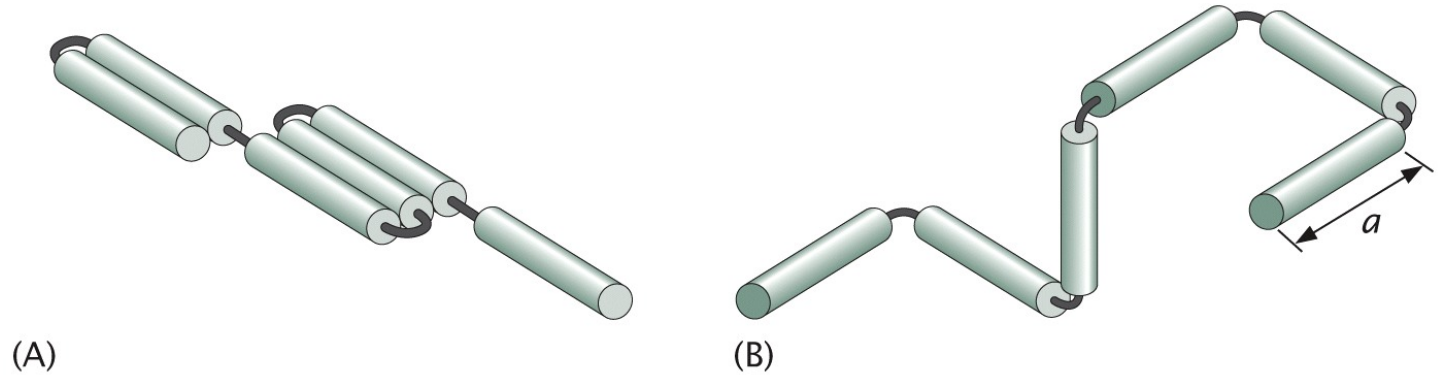


Figure 8.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

1D: segments at +/-180^0 with respect to each other

3D: where links are restricted to 90^0 angles

1D random walk model

The history of the walker is built as a sequence of N right and left moves, each with probability ½. Each move starts anew (no correlation).

- There are a total of 2^N admissible configurations
- All configurations are equally probable ($1/2^N$)

The mean size of the polymer scales as the square root of the number of segments ("diffusion")

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

The i-th step of the walk is $x_i = \pm a$

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

$$\langle R^2 \rangle = \left\langle \sum_{i,j=1}^{N} x_i x_j \right\rangle = \sum_{i=1}^{N} \langle x_i^2 \rangle + \sum_{i \neq j}^{N} \langle x_i x_j \rangle = Na^2$$

Probability of a given macromolecular *state*

Probability
of a single configuration

Degeneracy

$$p(n_r, N) = \frac{1}{2^N} W(n_r, N)$$

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

Probability
of a given *state*

$$\sum_{n_r=0}^{N} p(n_r, N) = 1$$

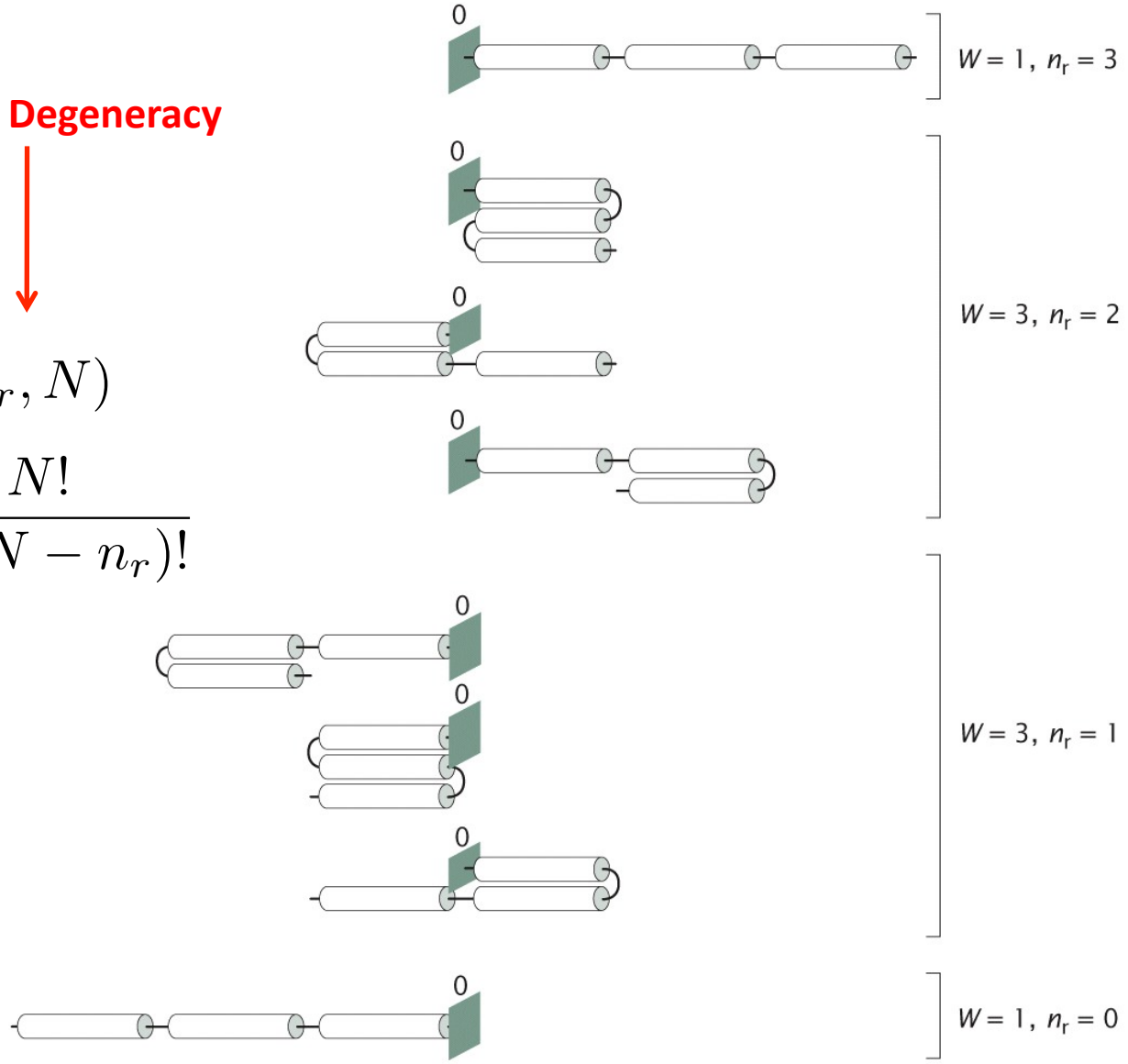
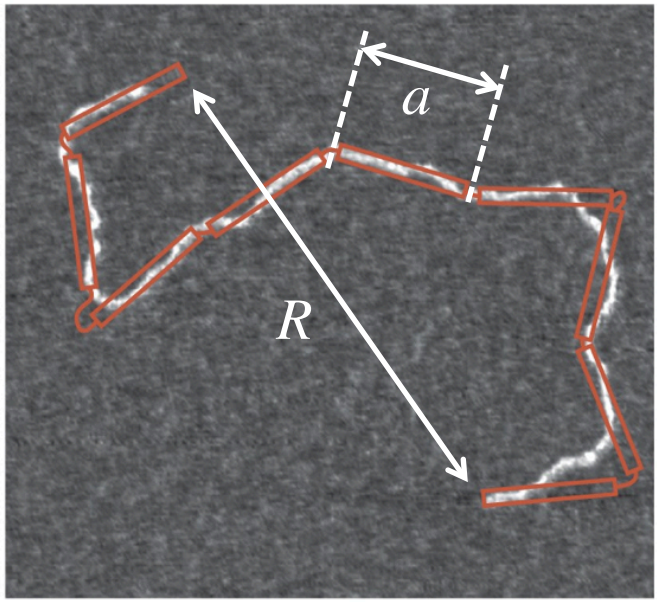


Figure 8.3 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Probability distribution of end-to-end distance

$$(n_r - n_l)a = R$$
$$n_r + n_l = N$$



$$p(n_r, N) \rightarrow p(R, N) = \frac{N!}{\left(\frac{N}{2} + \frac{R}{2a}\right)!\left(\frac{N}{2} - \frac{R}{2a}\right)!}\left(\frac{1}{2}\right)^N$$

Use Stirling approximation and expand the logarithms…

$$\log(N!) \approx N \log N - N + \log \sqrt{2\pi N}$$

$$\log(1 \pm x) = \pm x - \frac{x^2}{2} + \mathcal{O}(x^3)$$

After some algebra (do it!) → $\log p(R, N) = \log 2 - \log \sqrt{2\pi N} - \frac{R^2}{2Na^2}$

$$P(R, N) = \frac{p(r, N)}{2a} = \frac{1}{\sqrt{2\pi N a^2}} e^{-R^2/2Na^2}$$

Overwhelming probability of zero separation of ends

It is nothing but the **central limit theorem**

$$R = \sum_i x_i$$

is Gaussian-distributed, if xi are iid variables with finite mean and variance

… and it is Gaussian in three dimensions as well!

It immediately follows that the distribution of end-to-end separation is Gaussian in three dimensions as well (the conditions of the central limit theorem are met)

$$\langle \mathbf{R} \rangle = 0$$

$$\langle \mathbf{R} \cdot \mathbf{R} \rangle = Na^2$$

$$P(\mathbf{R}, N) = \mathcal{N} e^{-kR^2}$$

$$\int_{\mathbb{R}_3} P(\mathbf{R}, N)\, d^3R = 4\pi \int_0^\infty P(\mathbf{R}, N) R^2\, dR = 1$$

$$\int_{\mathbb{R}_3} R^2 P(\mathbf{R}, N)\, d^3R = 4\pi \int_0^\infty P(\mathbf{R}, N) R^4\, dR = Na^2$$

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

The persistence length

The persistence length is a measure of the length scale over which a polymer remains roughly straight

$\mathbf{t}(s)$ Tangent vector at a position *s* along the contour length

Persistence length

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

Average over all the configurations

This formula is valid for a free polymer. If the polymer is subject to constraints (i.e. ends attached) the correlation will change

Bacterial genome that has escaped the bacterial cell

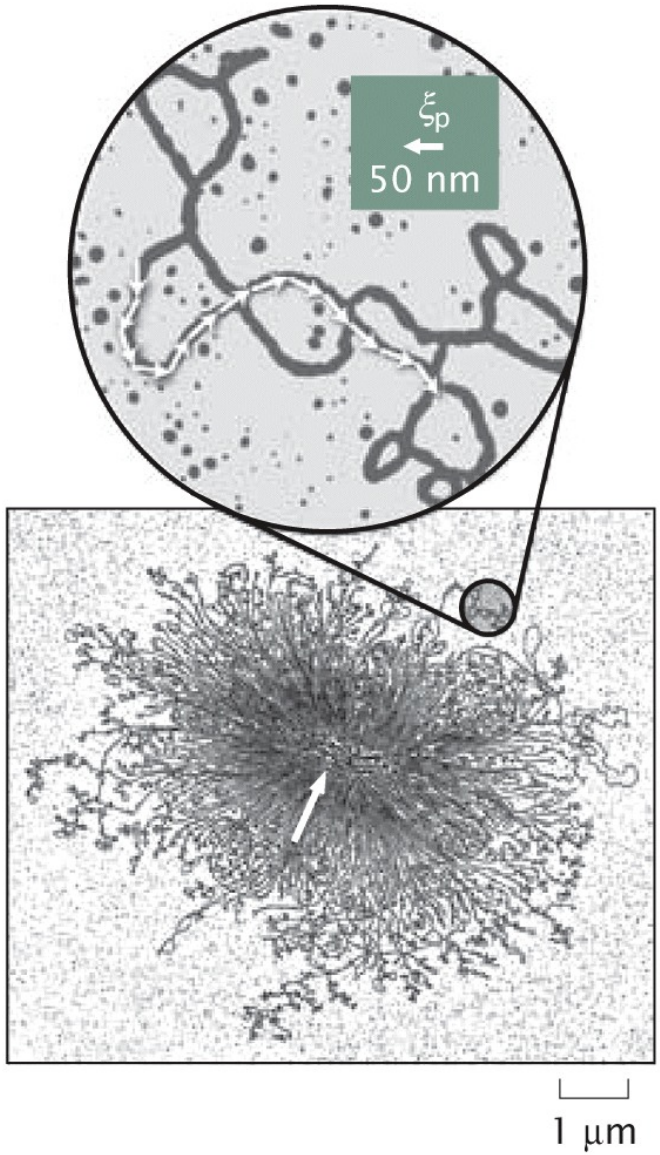


Figure 8.5 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The Kuhn length

A good example of long flexible polymer is provided by genomic DNA of viruses such as λ-phage, with a contour length of L = 16.6 μm.

This should be compared with the persistence length of DNA, ξ_p about 50 nm (room temperature and solvent conditions typical of the cellular environment).

By the very definition of persistence length, we can think of the **polymer as consisting of L / ξ_p connected links that take random orientations with respect to each other.** This is the logic that gives rise to the **freely-jointed chain** model

The length of the uncorrelated steps is known as **Kuhn length**

Let us derive the relation between the Kuhn length and the persistence length. We start from the end-to-end vector

$$\mathbf{R} = \sum_{i=1}^{N-1} (\mathbf{x}_{i+1} - \mathbf{x}_i) \Longrightarrow \int_0^L \mathbf{t}(s)\, ds$$

The Kuhn length and the persistence length

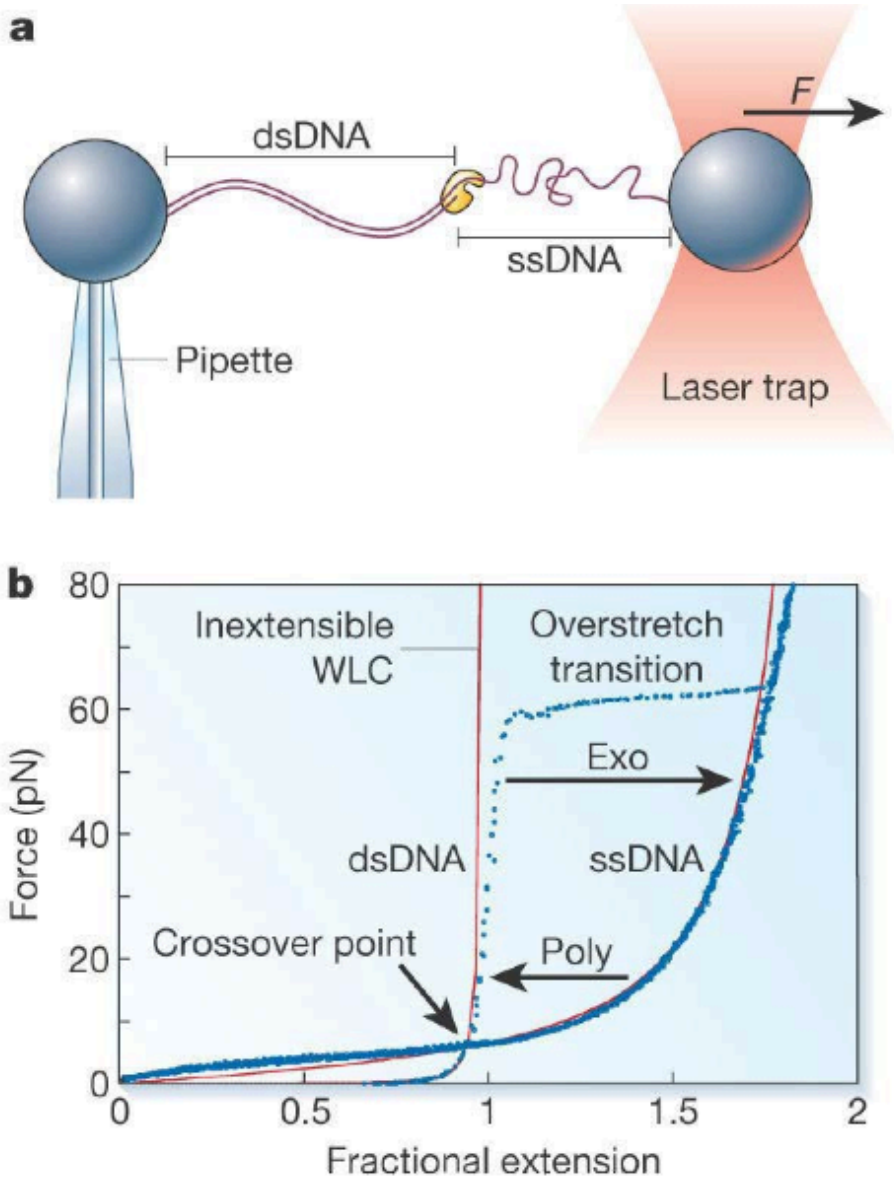
$$\langle \mathbf{R} \cdot \mathbf{R} \rangle = \int_0^L ds \int_0^L du \, \langle \mathbf{t}(s) \cdot \mathbf{t}(u) \rangle$$

$$= \int_0^L ds \int_0^L du \, e^{-|s-u|/\xi_p} \approx 2L\xi_p \qquad L \gg \xi_p$$

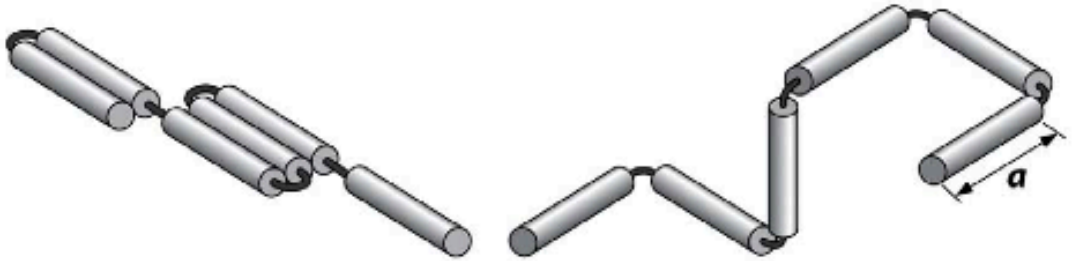
From the random walk model we know that $\langle \mathbf{R} \cdot \mathbf{R} \rangle = Na^2 = La \Longrightarrow a = 2\xi_p$

This gives a quantitative rule that allows to treat a given polymer with a given persistence length as a random walk (freely-jointed chain)

Interpreting force-extension curves of DNA

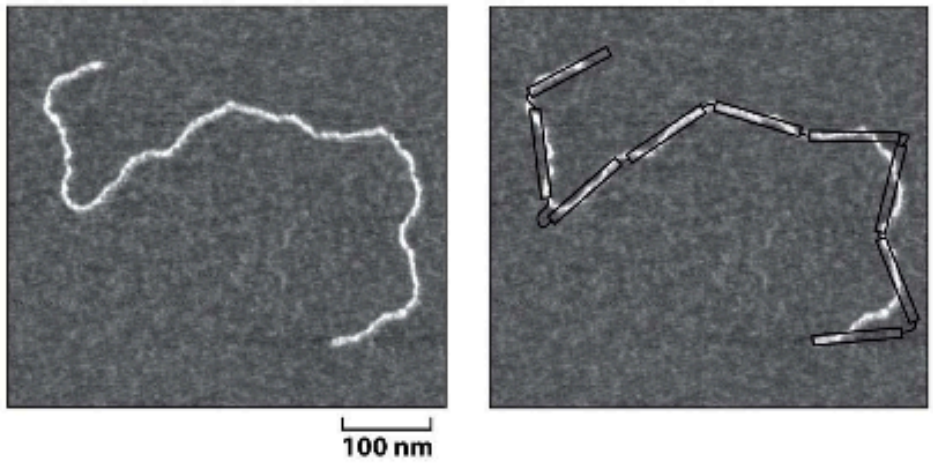


Freely-jointed chain model



One-dimensional Random Walk

Three-dimensional Random Walk



- Macromolecules are viewed as rigid segments connected by hinges (hypothetical).
- N bonds, N+1 rigid monomers
- All rigid segments have equal length.
- Every macromolecular configuration is equally probable.
- The molecule can sample many accessible configurations, all close in energy relative to kT.

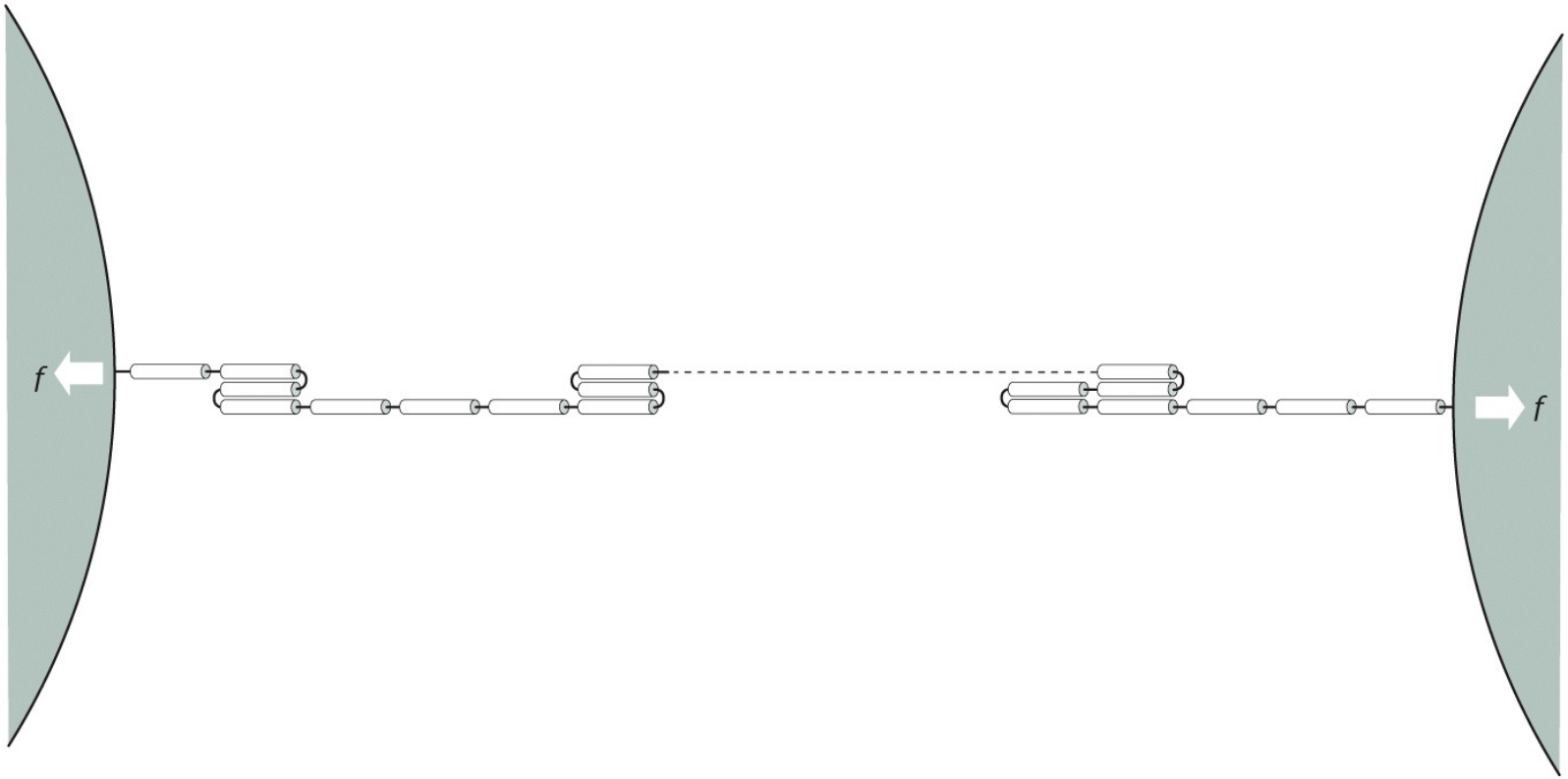


Figure 8.23 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The dangling mass analogy of a constant load

Figure 8.24 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Free energy in the presence of a load

$$G(L) = -fL - k_B T \log W(L, N)$$

As the chain stretches the "mass" is lowered and the potential energy decreases

The more conformations available for a given contour length L, the higher the entropy and the lower the free energy

$$W(n_r, N) = \frac{N!}{n_r!(N - n_r)!} \qquad L = (n_r - n_l)a$$

Using Stirling approximation and neglecting constant terms (only free-energy differences matter!)

$$G(L) = -2fa\, n_r + k_B T[n_r \log n_r + (N - n_r) \log(N - n_r)]$$

The most probable value of *L* minimizes the free energy

$$\frac{\partial G(L)}{\partial n_r} = 0 \Longrightarrow \begin{cases} n_r = \dfrac{N}{1 + e^{-2fa/k_BT}} \\ n_l = \dfrac{Ne^{-2fa/k_BT}}{1 + e^{-2fa/k_BT}} \end{cases}$$

$$L = (n_r - n_l)a = Na \tanh\left(\frac{fa}{k_BT}\right)$$

In the small-force regime, the DNA behaves as an *entropic spring*

$$fa \ll k_BT \Longrightarrow \begin{cases} \text{Double-stranded DNA, } a \approx 100 \text{ nm} \Longrightarrow f \ll 40 \text{ fN} \\ \text{Single-stranded DNA, } a \approx 1.5 \text{ nm} \Longrightarrow f \ll 3 \text{ pN} \end{cases}$$

In three dimensions …

The calculations are simpler if we assume that at each site the monomer vector can assume all orientations.

The one-particle partition function

$$
\begin{aligned}
Z_1 &= \int_0^{2\pi} d\varphi \int_0^{\pi} e^{f a \cos\vartheta / k_B T} \sin\vartheta \, d\vartheta \\
&= 2\pi \int_{-1}^{1} e^{f a \, x / k_B T} \, dx \\
&= 4\pi \left(\frac{k_B T}{f a} \right) \sinh \left(\frac{f a}{k_B T} \right)
\end{aligned}
$$

The N-particle partition function

$$Z = (Z_1)^N$$

Note that the factor 1/*N*! is absent here. The reason is that the monomers are not undistinguishable and the polymer is virtually *directed* (ends fixed)

$$G = -k_B T \log Z = -k_B T \left\{ \log \left[\sinh \left(\frac{fa}{k_B T} \right) \right] - \log \left[\frac{fa}{k_B T} \right] \right\}$$

The average chain length is the thermodynamically conjugate variable of the external force *f*

$$\langle L \rangle = -\frac{\partial G}{\partial f} = Na \left[\coth \left(\frac{fa}{k_B T} \right) - \frac{k_B T}{fa} \right]$$

$$\coth \simeq \frac{x}{3} + \frac{1}{x} \quad \text{as } x \to 0$$

$$\frac{\langle L \rangle}{N} \simeq \left(\frac{a}{3 k_B T} \right) f$$

Again *entropic spring*. The force constant is three times stiffer than in 1*D*

$$\frac{\langle L \rangle}{L_{\text{tot}}} = \begin{cases} \tanh\left(\frac{fa}{k_B T}\right) \simeq \left(\frac{a}{k_B T}\right) f & 1D \\ \coth\left(\frac{fa}{k_B T}\right) - \frac{k_B T}{fa} \simeq \left(\frac{a}{3k_B T}\right) f & 3D \end{cases}$$

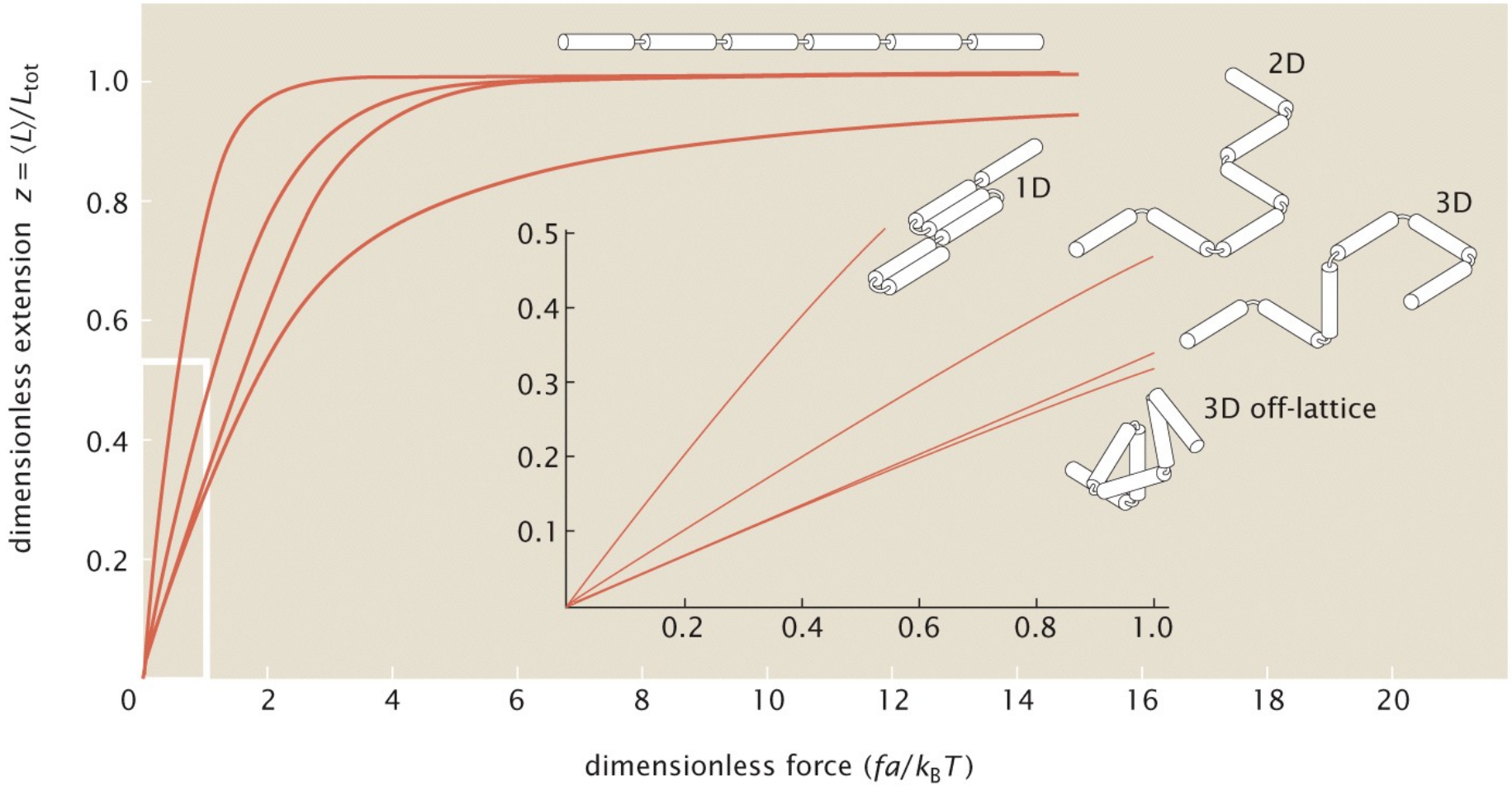


Figure 8.25 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

What do the experiment reveal?

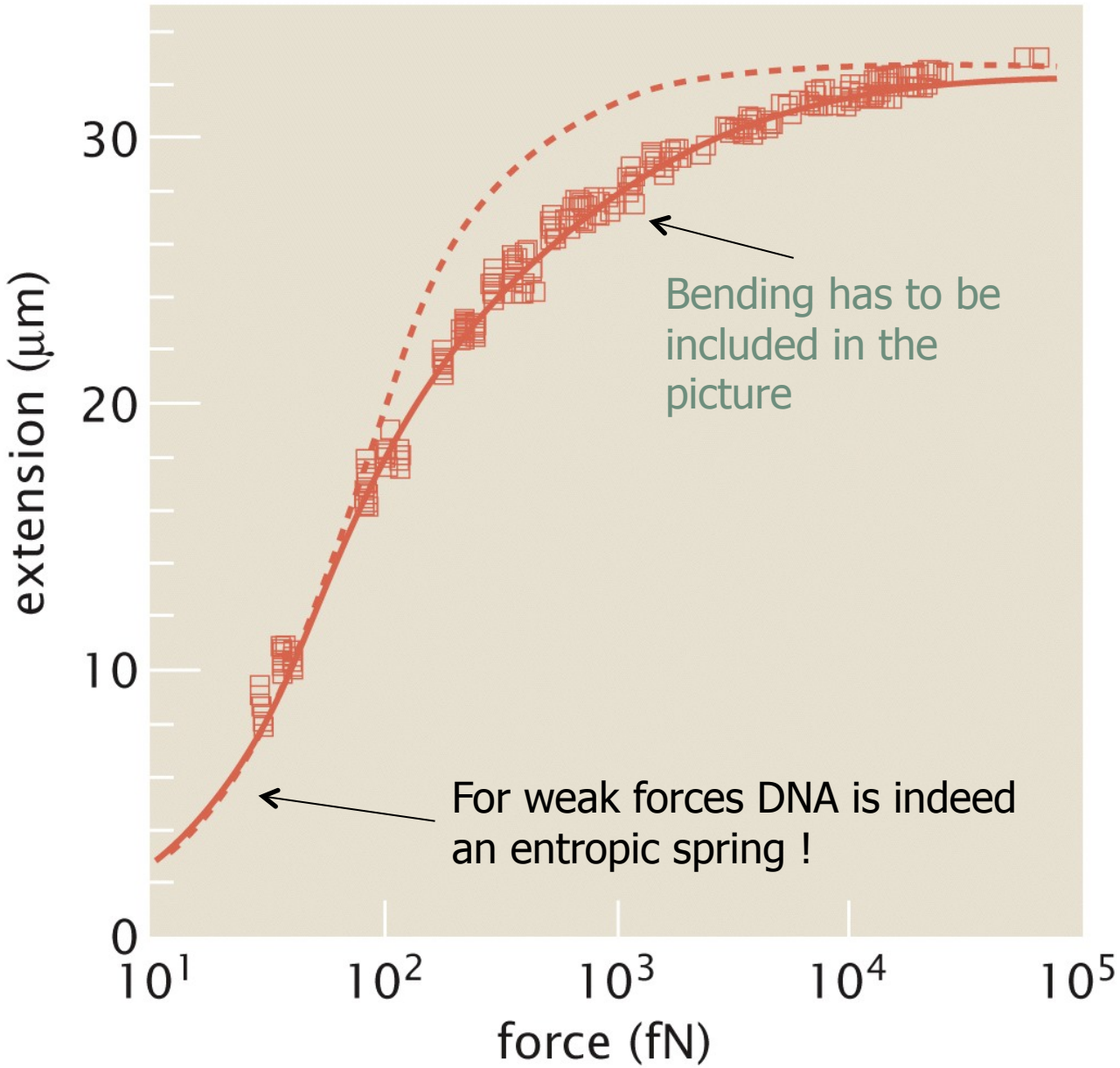


Figure 5.14 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Dashed line:
Freely jointed chain

Solid line:
Worm-like chain

Adapted from
Bustamante et al.
Science 265, 1599 (1994)