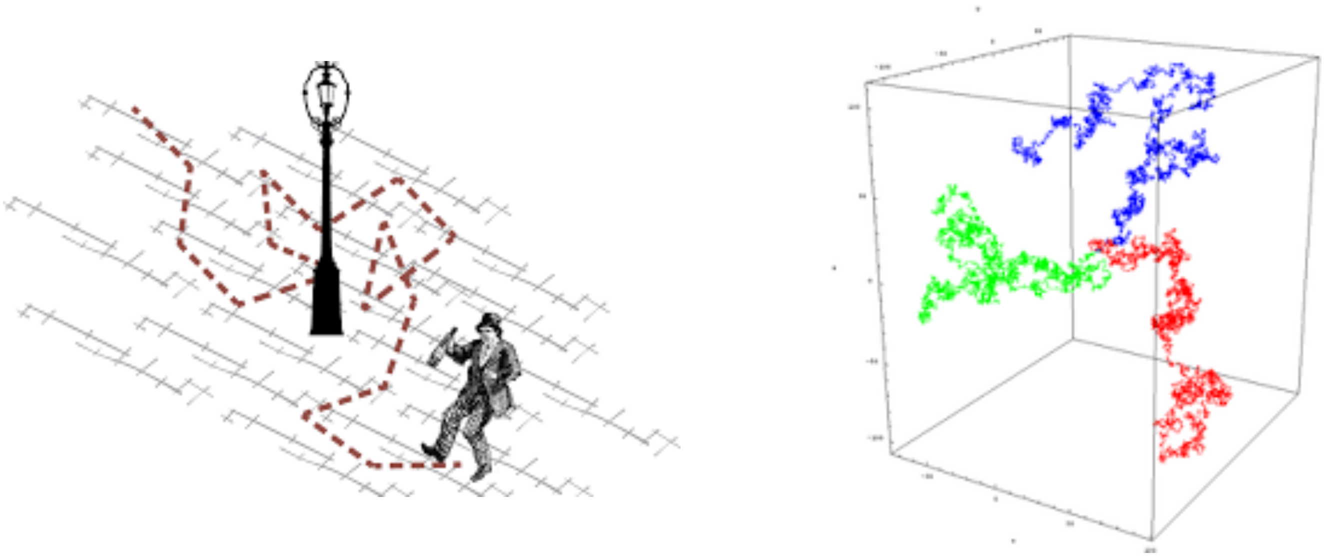
Lecture of November 21: Chapter 8 Model of DNA and Proteins

Apichart Linhananta
Department of Physics
Lakehead University

Random Walk

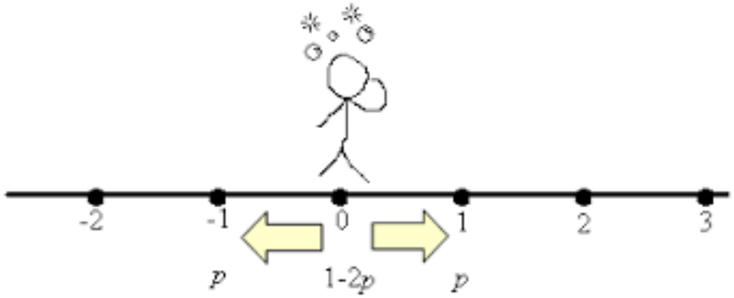
- Stochastic means having a probability distribution that may be analyzed statistically, but cannot be predicted precisely.
- A Random Walk is a Stochastic Path determined by steps whose length and direction is determined by a probability distribution.
- For large number steps the distribution is Gaussian. This is the central limit theorem.

Random Walk



- On average, after a large number of steps, the random walker position will be, $\langle x \rangle = 0$ (1D), $\langle \vec{r} \rangle = 0$ (3D)
- After a large number of steps, N, the random walker will end up far away from where he started: $\langle x^2 \rangle = NL^2$ (1D), $\langle R^2 \rangle = 2NL^2$ (3D), where L is the average step size.

1D Random Walk: Step Size a; N steps



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

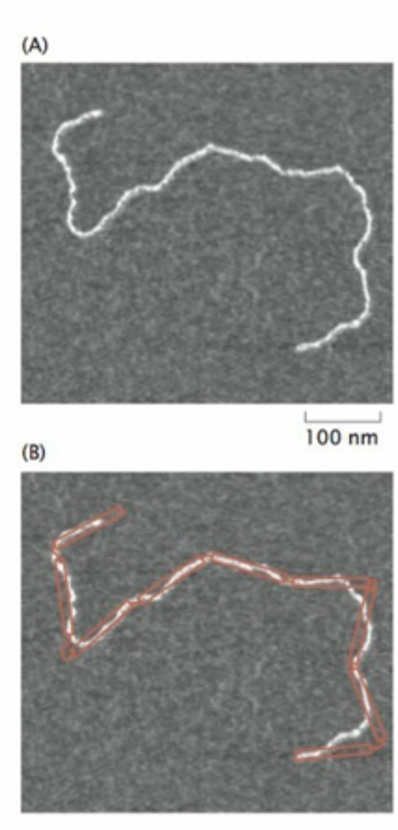
$$\langle R^2 \rangle = N a^2 .$$

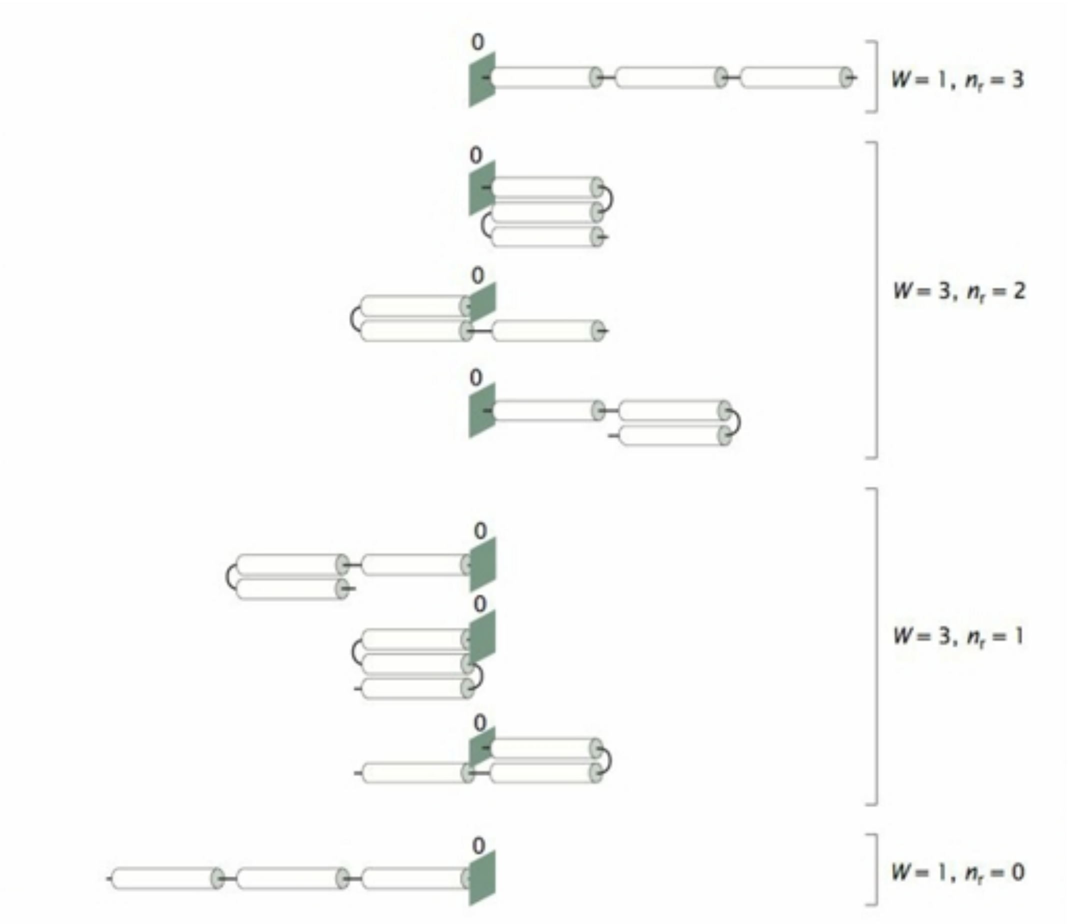
See Class Notes for Detail

DNA, RNA and Proteins as a Random Walk

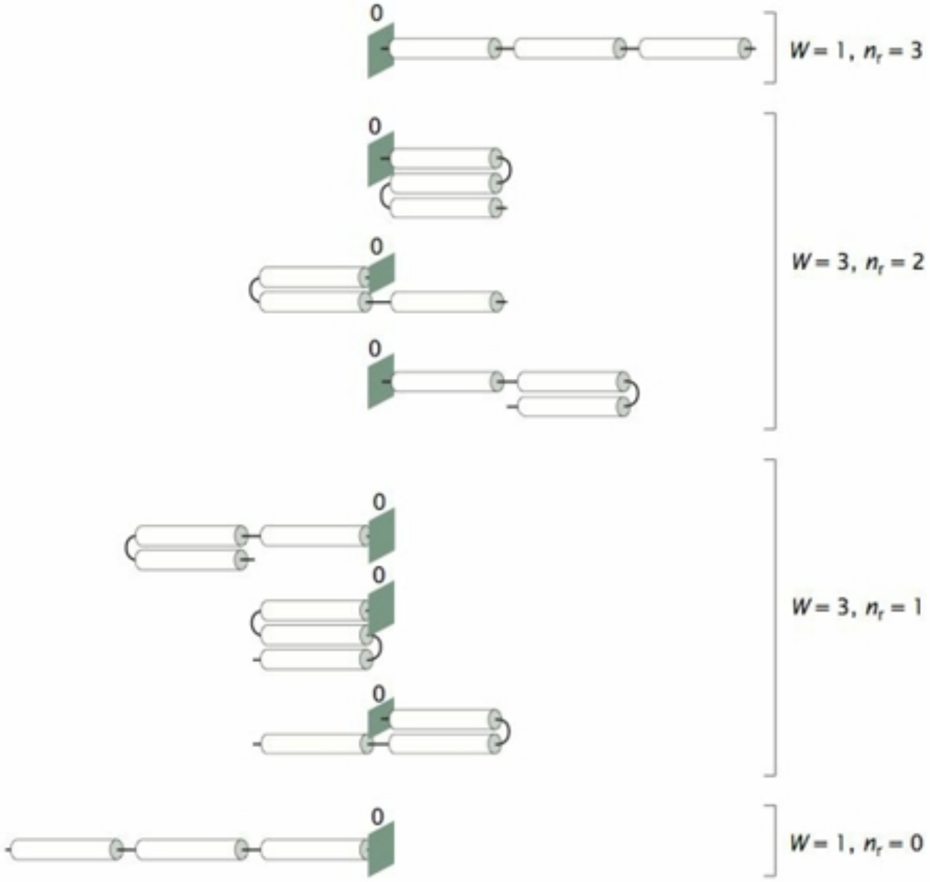
- Long molecules made of repeating units are called polymers.
- Homopolymers are polymers made up of one type of unit. An example is polyethylene (plastic)
- DNA, RNA and proteins are heteropolymers since they are made up of type of units (i.e. nucleotides, amino acids

DNA conformations as a random walk





DNA conformations as a random walk



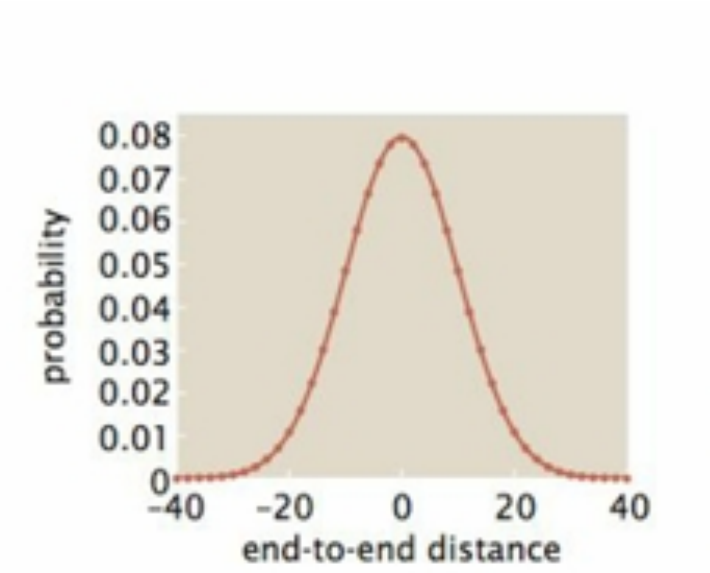
$$W(n_r; N) = \frac{N!}{n_r!(N - n_r)!}, \qquad (8.5)$$

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

DNA end-to-end distance is a Gaussian Distribution

$$\sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} = 2^N.$$

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



$$p(R;N) = \frac{2}{\sqrt{2\pi N}} e^{-R^2/2Na^2}.$$

1D Discrete probability

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

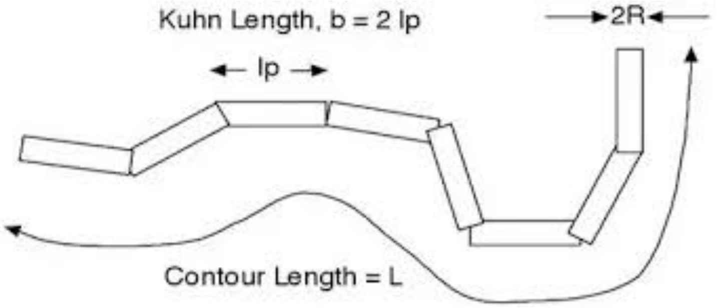
1D continuous probability per unit length

3D Distribution and Kuhn Length

$$P(\mathbf{R}; N) = \left(\frac{3}{2\pi N a^2}\right)^{3/2} e^{-3R^2/2Na^2}. \qquad (8.23)$$

a is the **Kuhn Length or Segments**

- A Kuhn segment is considered to be a rigid. A segment on the polymer/DNA that is much longer than the Kuhn length ($\gg a$) is considered to be flexible.
- 2 adjacent Kuhn segments can be thought of as being **freely jointed** to each other.



3D Distribution and Kuhn Length

$$P(\mathbf{R}; N) = \left(\frac{3}{2\pi Na^2}\right)^{3/2} e^{-3R^2/2Na^2}. \qquad (8.23)$$

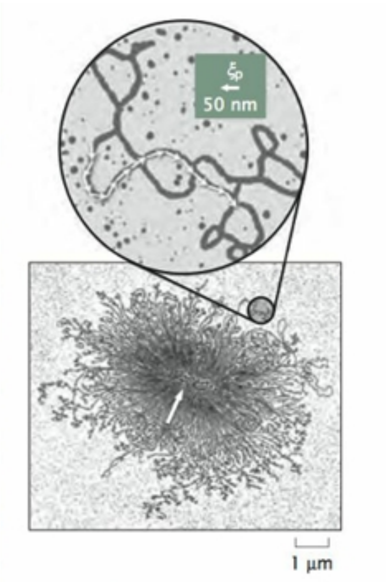
a is the **Kuhn Length or Segments**

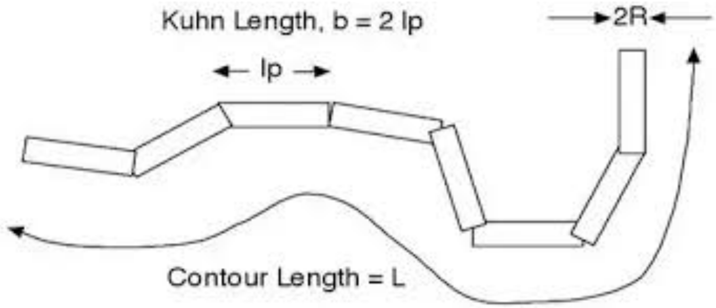
- For DNA a Kuhn segment is about 300 bp.

End-to-end distance (size)

$$R_{ee}^2 = Na^2 \rightarrow R_{ee} = \sqrt{N}a$$

Above confirms figure below for size of **unconfined** E. Coli's DNA $\sim 4.5 \times 10^6 bp$



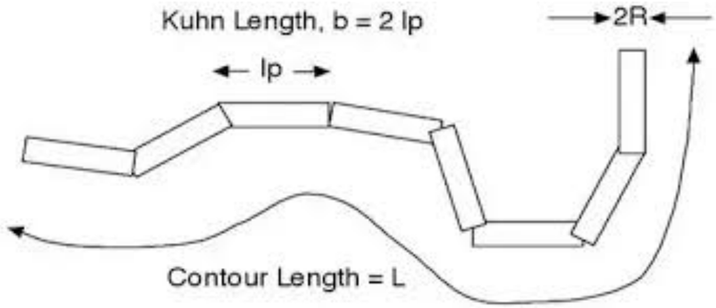


3D Distribution and Kuhn Length and Persistent Length

ξ_p is the **Persistent Length** is the length over which the polymer (DNA) can be considered straight.

- In polymer physics, it is usually assumed that the persistent length is half the Kuhn length.

$$\xi_p = a/2$$



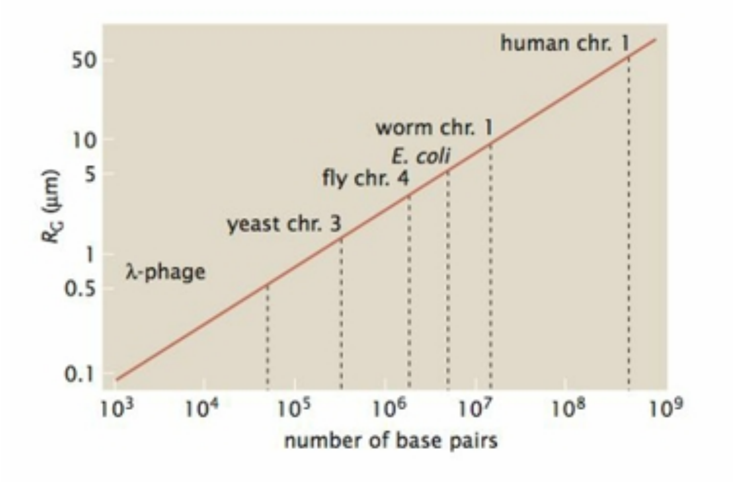
3D Distribution and Radius of Gyration

A better measure of the size of a DNA is the **Radius of Gyration**,

$$R_G = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\vec{R}_i - \vec{R}_{CM}\right)^2} = \sqrt{\frac{L\xi_p}{3}}$$

In comparison the **end-to-end** distance is

$$R_{ee} = \sqrt{2L\xi_p}$$



Section 8.2.3: Geography of Chromosomes

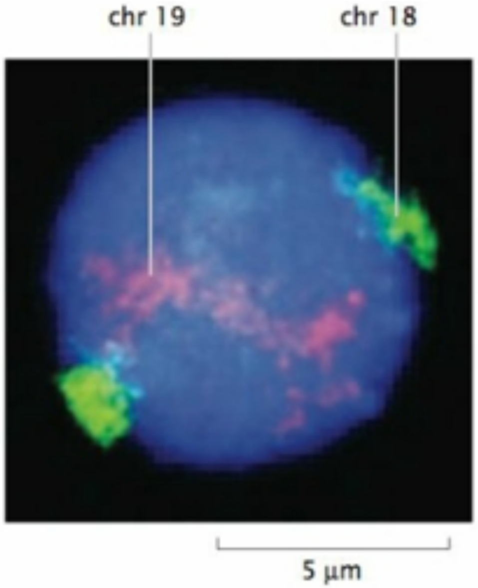


Figure 8.8: Fluorescently stained chromosomes 18 (green) and 19 (red) in the nucleus of a human cell. The two copies of chromosome 18 typically assume positions near the periphery of the nucleus, while the two copies of chromosome 19 are located closer to the center. (From J. A. Croft et al., *J. Cell Biol*. 145:1119, 1999.)

Tethered Chromosomes

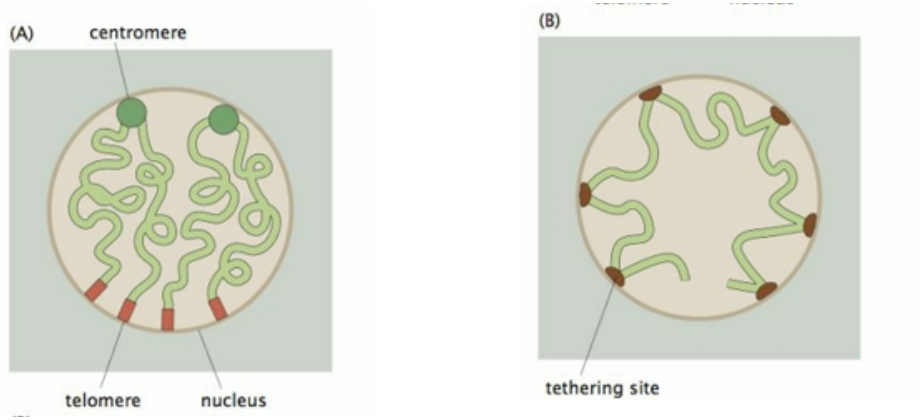


Figure 8.9: Cartoon representation of possible tethering scenarios of interphase chromosomes. (A) Tethering at the centromere and the two telomeres at the nuclear periphery and (B) tethering at intermediate locations. (Adapted from W. F. Marshall, *Curr. Biol.* 12:R185, 2002.)

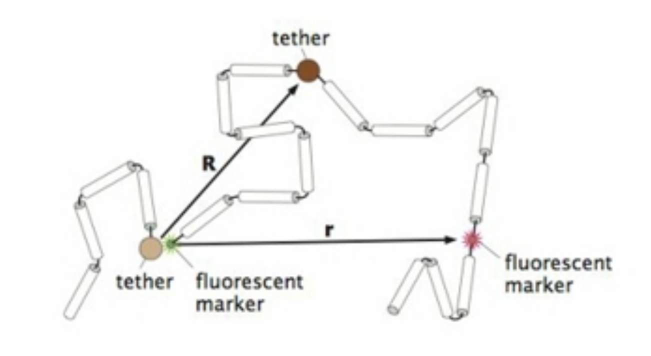
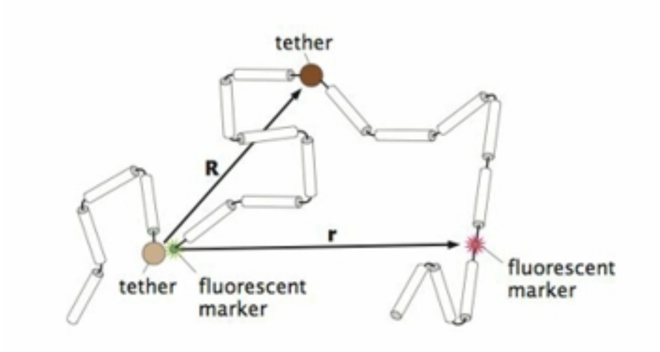


Figure 8.10: Simple configuration of a tethered chromosome. The two tethers are at fixed locations in space, and the second tether is at position **R** with respect to the first. The distribution of distances between the two fluorescent markers, one being at the same position on the chromosome as the tether, is a displaced Gaussian.

Tethered Chromosomes



$$P(r) = \left(\frac{3}{2\pi N' a^2}\right)^{1/2} \frac{r}{R}\left(\mathrm{e}^{-3(r-R)^2/2N'a^2} - \mathrm{e}^{-3(r+R)^2/2N'a^2}\right) \qquad (8.37)$$

Tethered Polymer

$$P(r) = \left(\frac{3}{2\pi N a^2}\right)^{3/2} 4\pi r^2 \mathrm{e}^{-3r^2/2Na^2}, \qquad (8.36)$$

Free Polymer

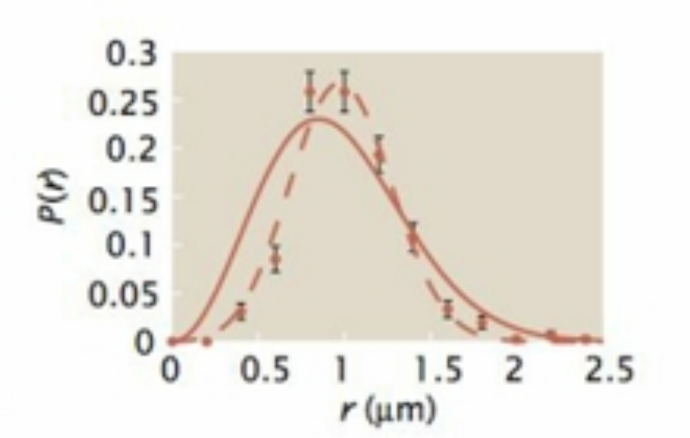


Figure 8.12: Statistics of yeast chromosome III. Distribution of distances between two fluorescent tags placed in proximity of the centromere and the HML region on yeast chromosome III. These two regions are separated by approximately 100 kb in genomic distance. The full line is a fit to the free-polymer distance distribution, Equation 8.36, while the dashed line is a fit to the tethered-polymer formula, Equation 8.37. (Courtesy of S. Gordon-Messer, J. Haber, and D. Bressan.)

Confined DNA: Chromosomes

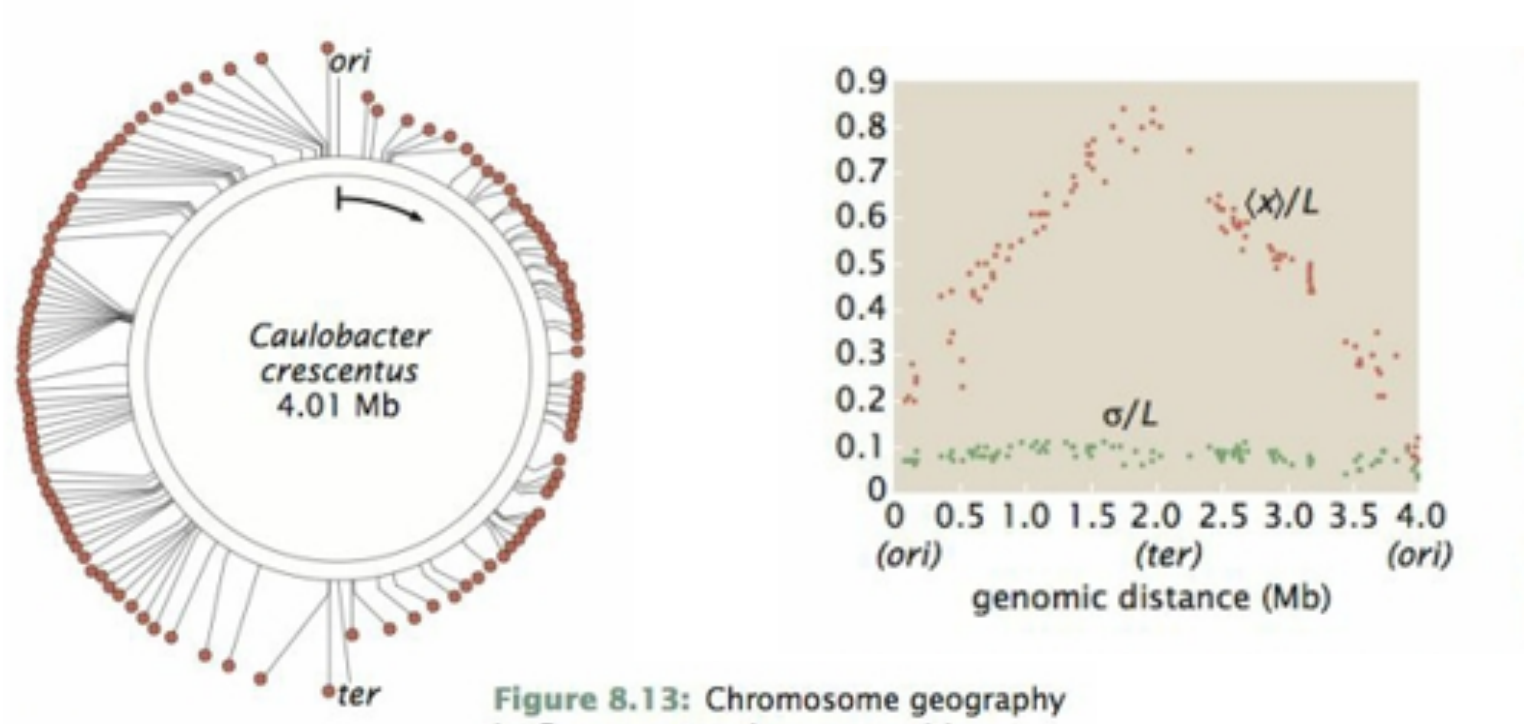
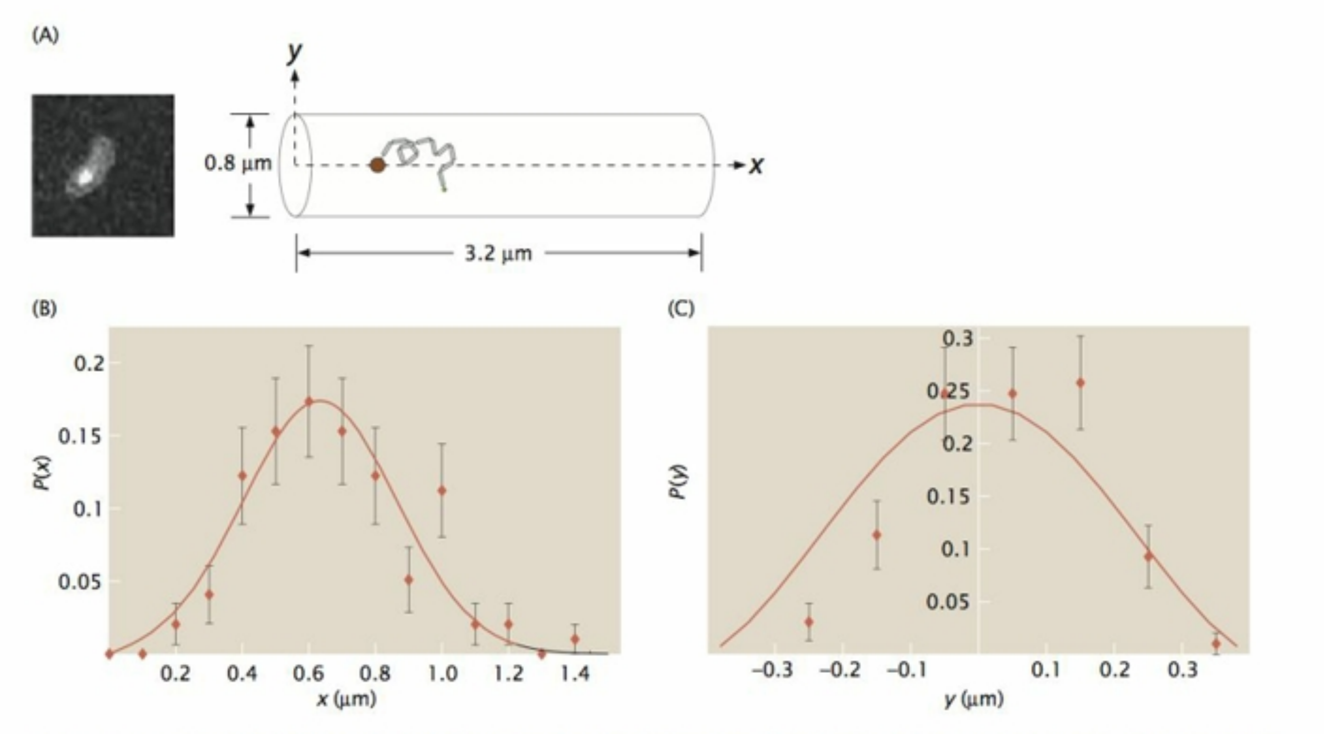


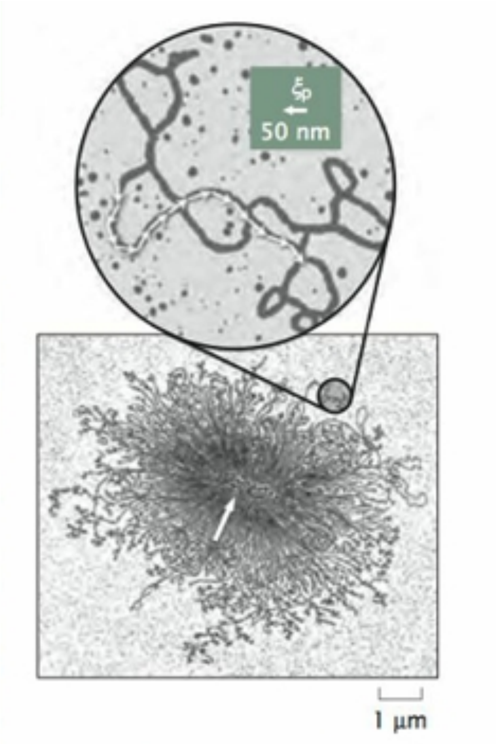
Figure 8.13: Chromosome geography in *C. crescentus*. Average positions (x/L) and the standard deviation ($\Delta x/L$) of the position along the long axis of the cell are plotted for 112 different fluorescently tagged locations along the chromosome of *C. crescentus*. The mean locations of the fluorescent tags and their corresponding standard deviations are shown on the diagram. (Adapted from P. H. Viollier et al., *Proc. Natl Acad. Sci. USA* 101:9257, 2004.)

Confined DNA: *V. cholerae* Chromosomes

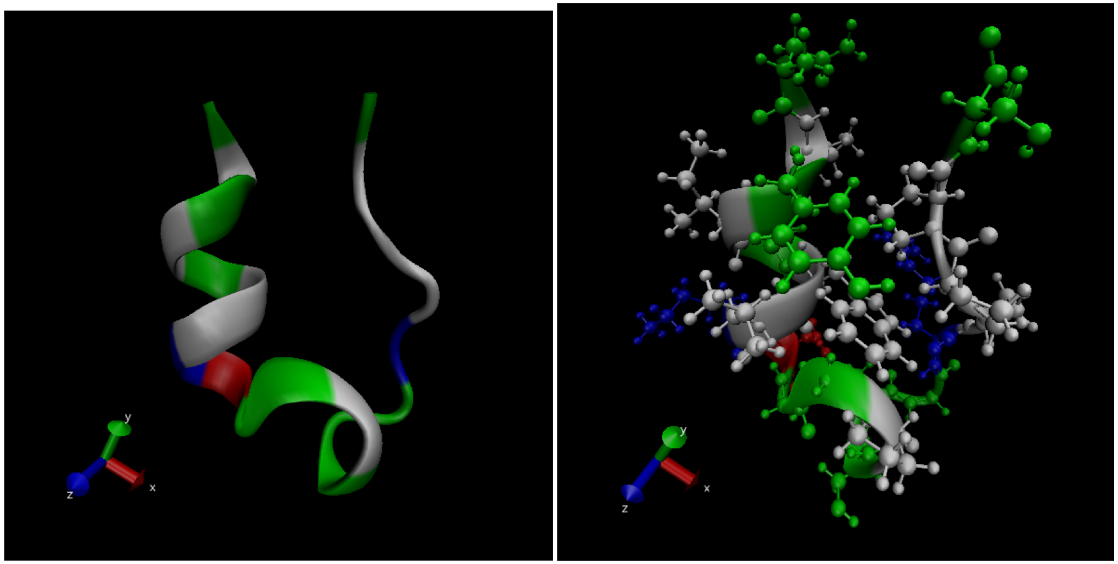


Gaussian in the x-direction (along cylinder axis), Confined Gaussian in the radial direction.

Comparisons of Proteins and DNAs

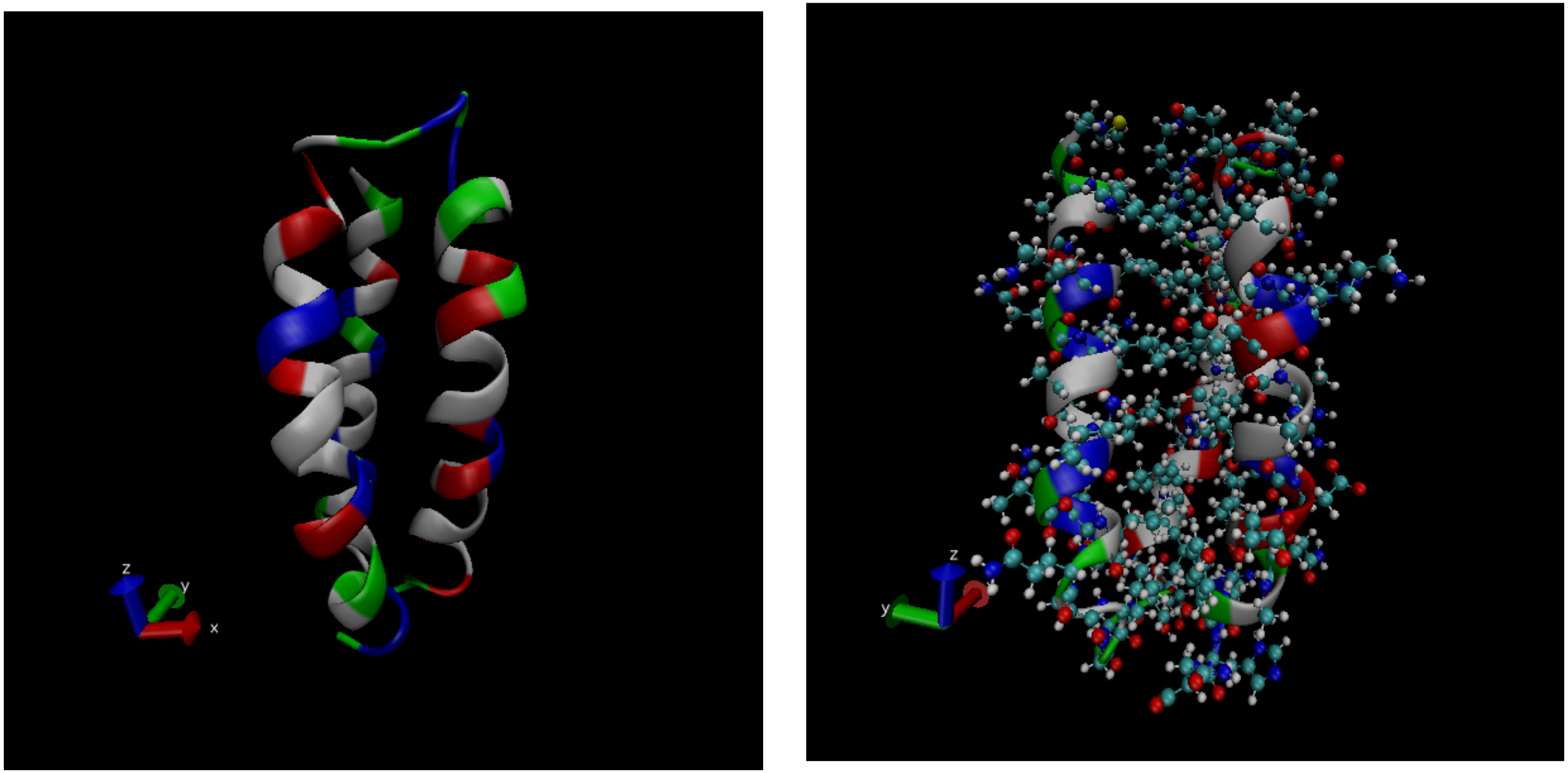


Unconfined DNA are unstructured Random-Walk Polymer



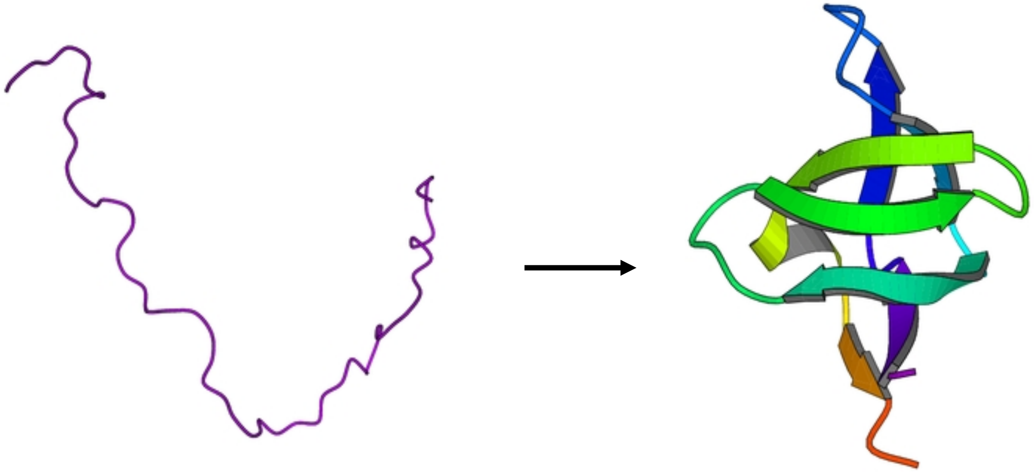
Ribbon and CPK representations of Trpcage Protein shows a compact (and unique) **Native** structure, with hydrophobic (nonpolar) moiety in the core, in gray or white.

Three-Helix-Bundle Protein Native Structure



Ribbon and CPK representations of Three-Helix-Bundle Protein shows a compact (and unique) structure, with hydrophobic (nonpolar) moiety (in gray or light blue) in the core.

Proteins:Levinthal Paradox (1968)

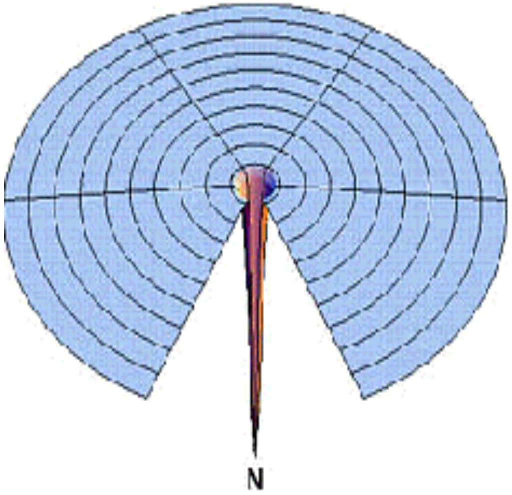


- 100 residues
- 2 conformation/residue
- 10^{30} conformations
- 10^{-12} s to convert

Random search to unique native state

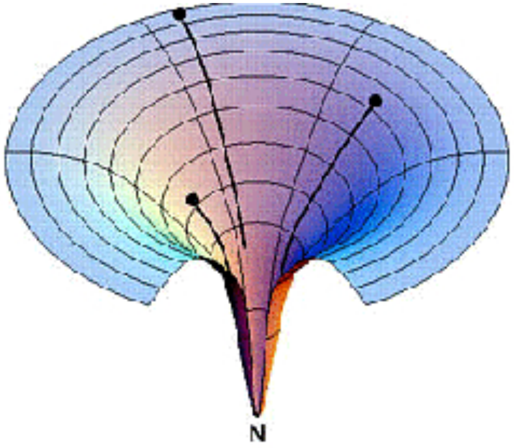
$\sim 10^{-12} \text{x } 10^{30} = 10^{18}\text{ s} \sim 10^{11}$ years

Proteins:Folding Free-Energy Landscape



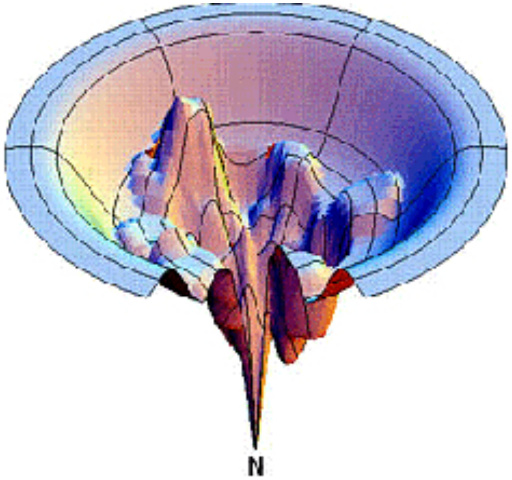
Uniform landscape

- Levinthal paradox



Funnel landscape

- Guided Pathways

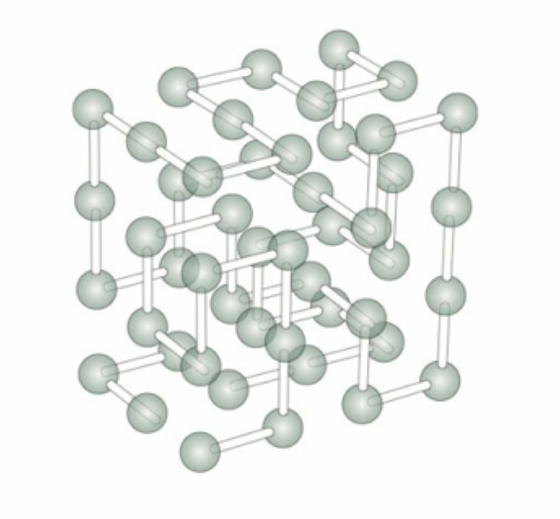


Rough landscape

- Funnel-like
- Traps

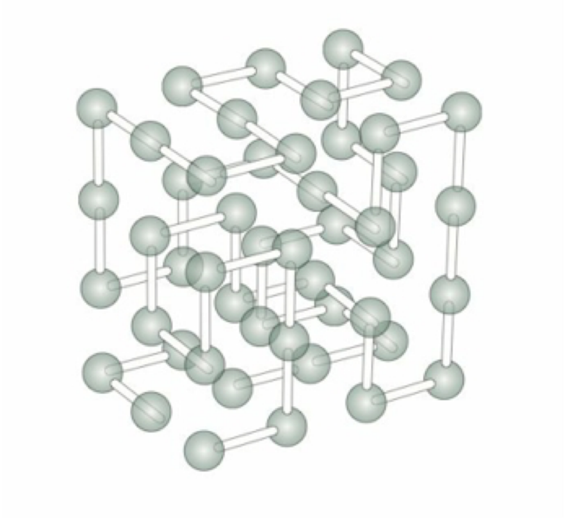
- Dill and Chan, **Nature Struct. Biol. 4**, 10 (1997)

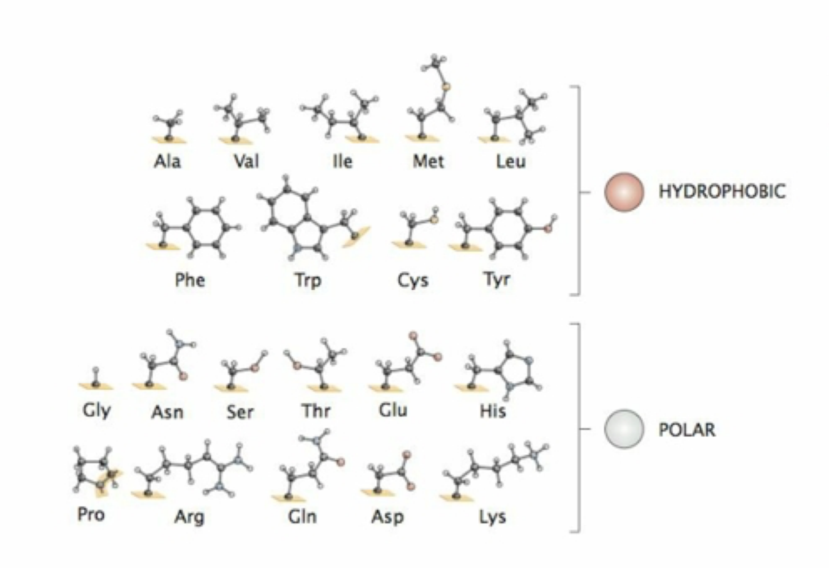
Simple Model of Protein Folding: HP lattice Model of Dill and Chan



One amino acid (aa) occupy one lattice and interact with each other.

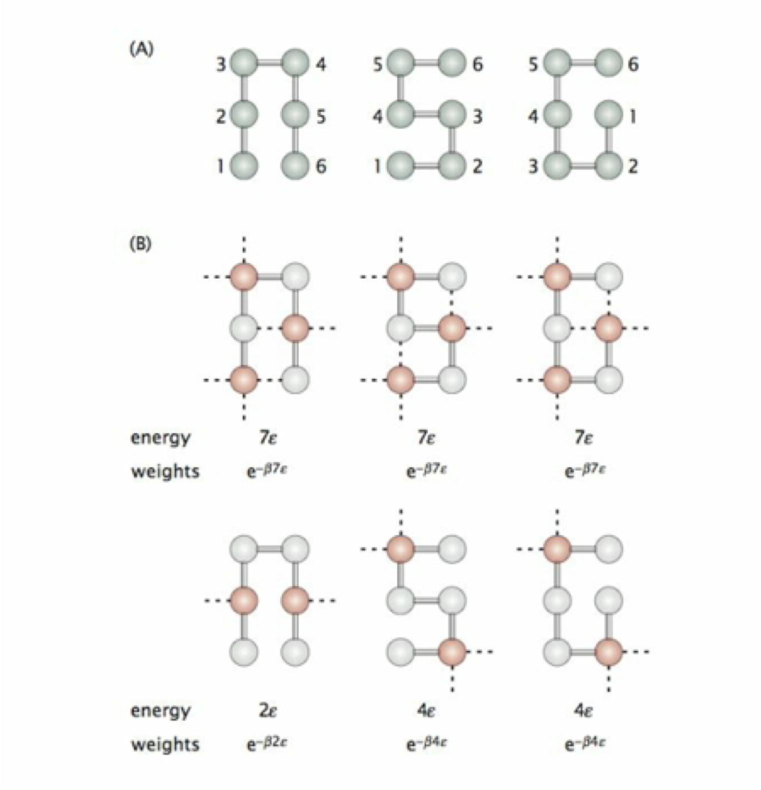
Simple Model of Protein Folding: HP lattice Model of Dill and Chan

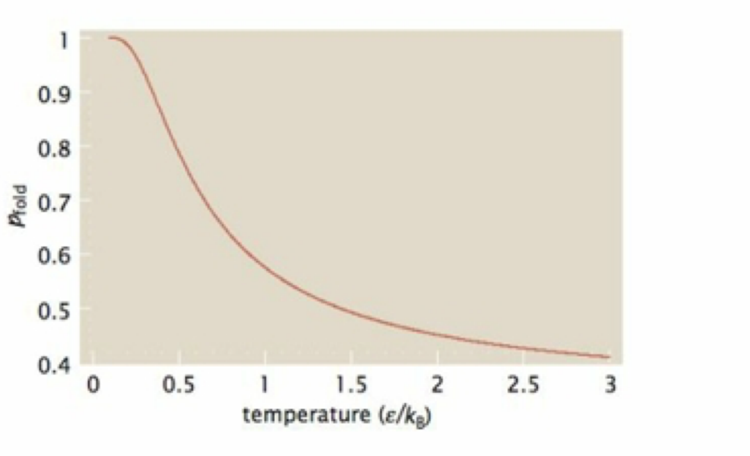




- In HP model there are only two kinds of aa: Hydrophobic (H); Polar (P)
- HH and PP interactions are more favorable that HP interaction.

Simple Model of Protein Folding: HP lattice Model of Dill and Chan





- Folded Native (F) state has lowest energy 2ε.
- At very low temperature the protein is folded.