

2015 Summer School on Polymers in Biology

DNA mechanics and structural diversity of DNA

@ KIAS, 22 Jun – 3 July
Seok-Cheol Hong
Korea University

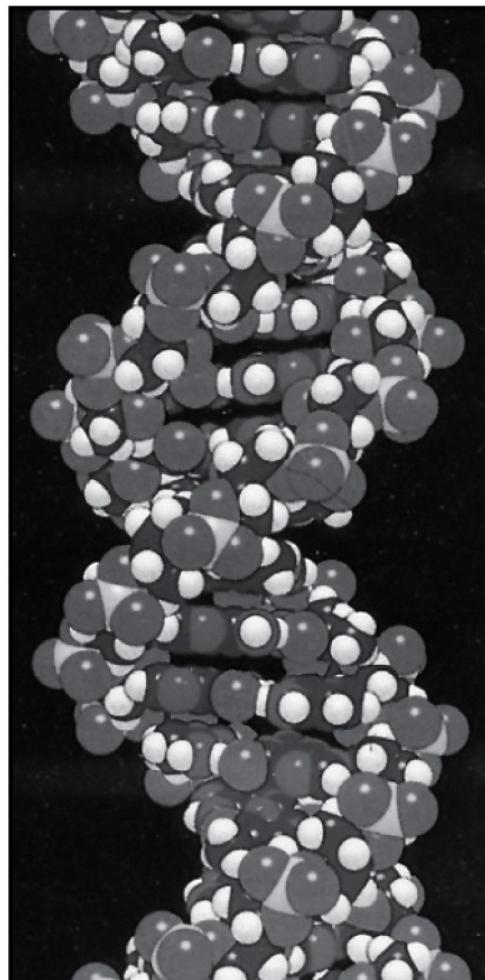
Lecture 1

- Hierarchy of biological organization
- Biomolecules: 1D polymers
- Examples of Polymers in Biology: DNA, RNA, Proteins, and Polysaccharides
- DNA: genetic material; double helix
- Central Dogma
- DNA thermodynamics
-

Lecture 2

- Watson-Crick base pair
- Effects of chemical factors on DNA stability
- DNA sequence vs. charged polymer
- Mechanical models: Freely Jointed Chain model
- Persistence length, end-to-end extension, radius of gyration, force response
-

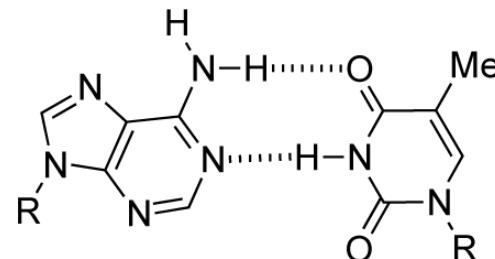
Watson-Crick base pair: canonical



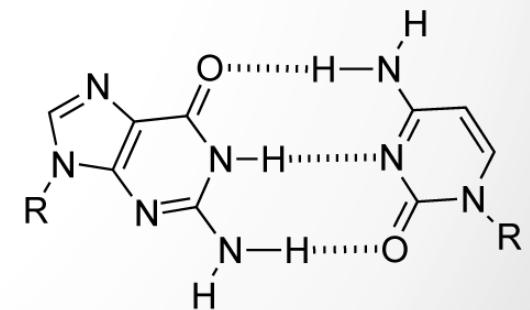
0.34 nm

minor
groove

major
groove



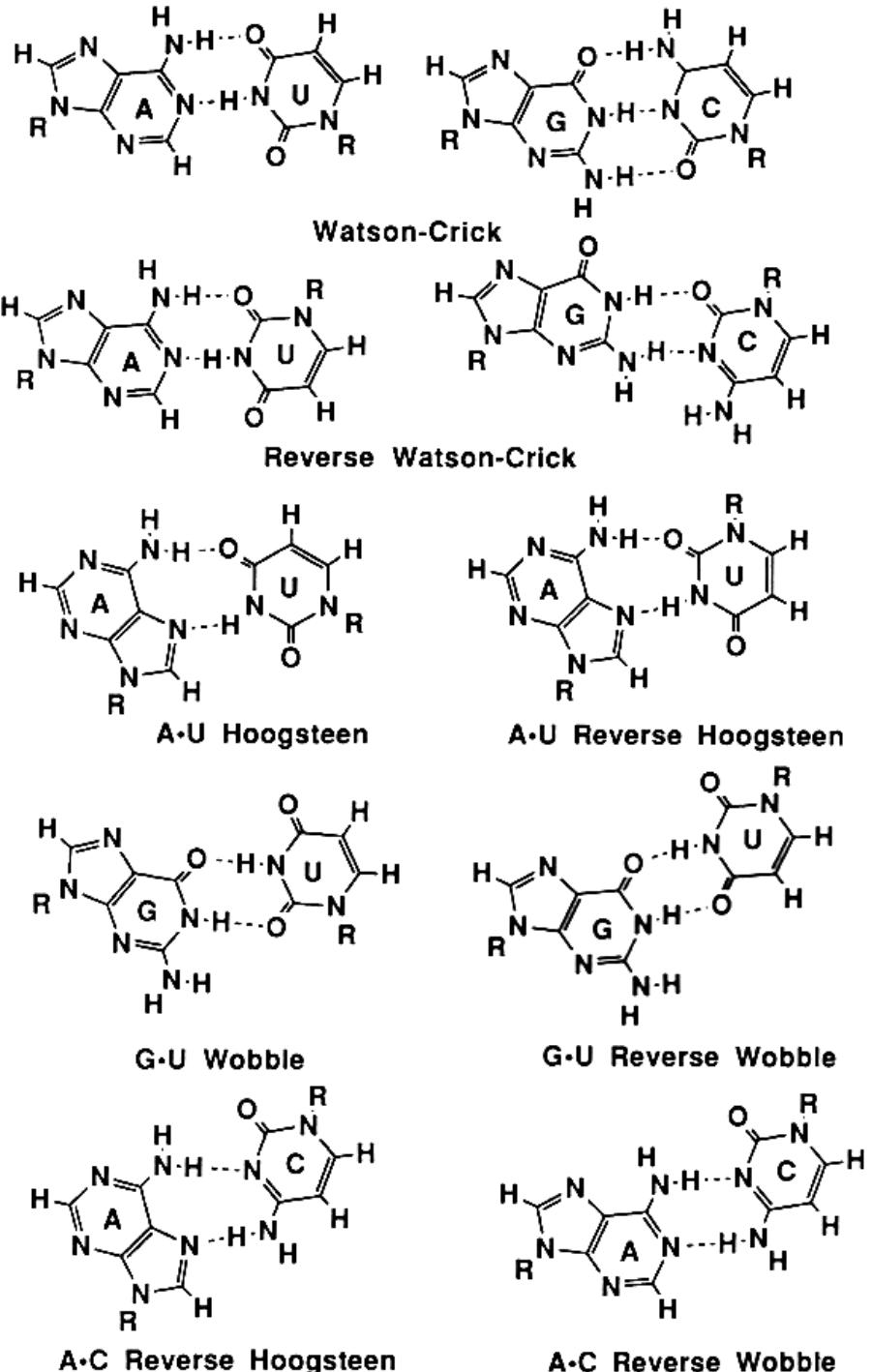
A-T base pair



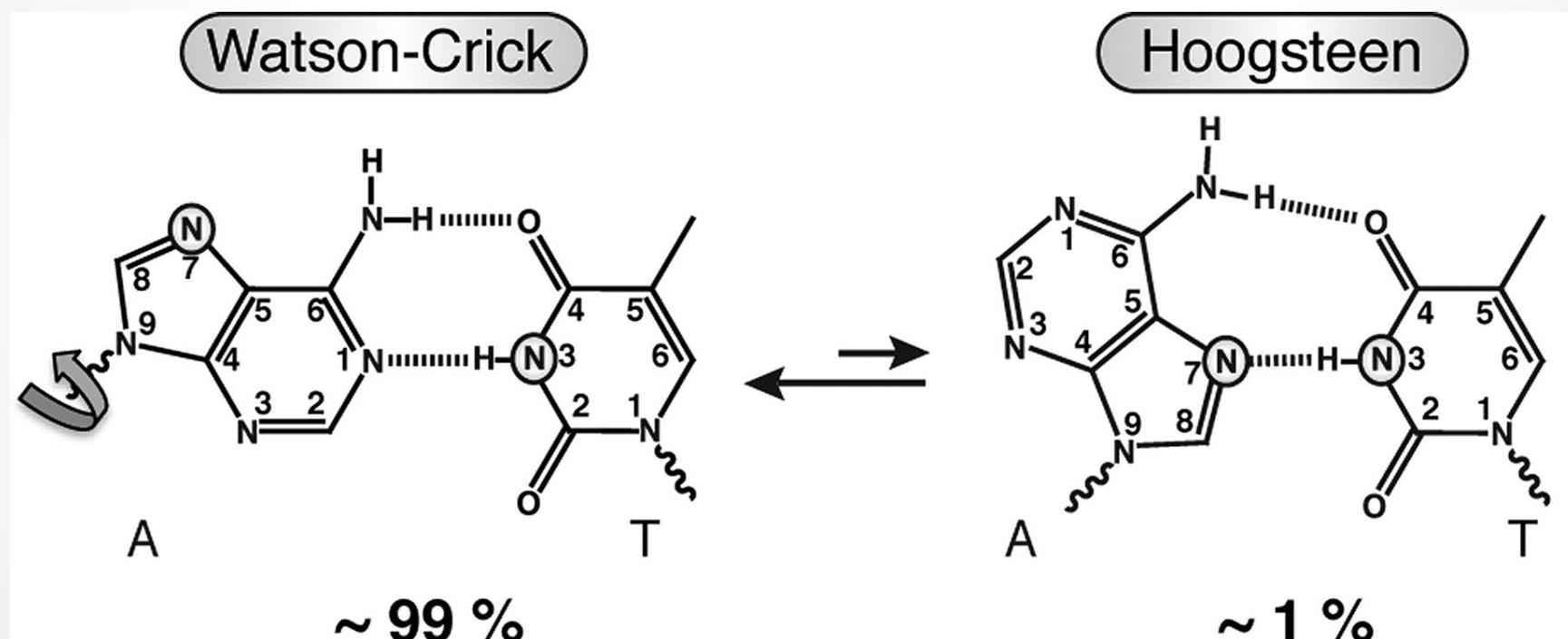
G-C base pair

Non-canonical base pairs

- Hoogsteen
- Wobble
- Reverse Hoogsteen
- Reverse Wobble

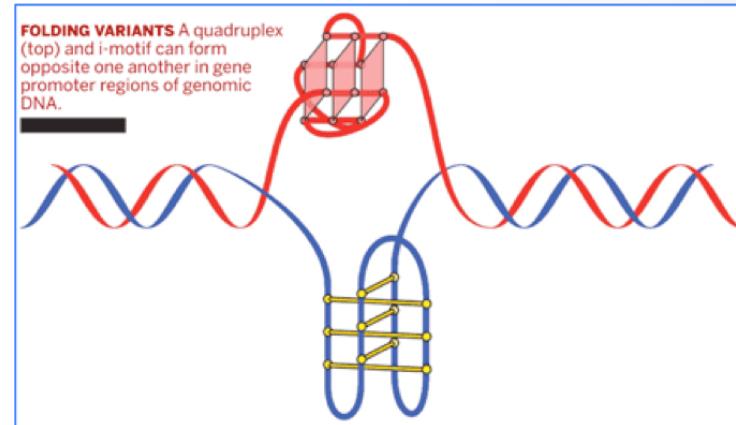


Hoogsteen base pair

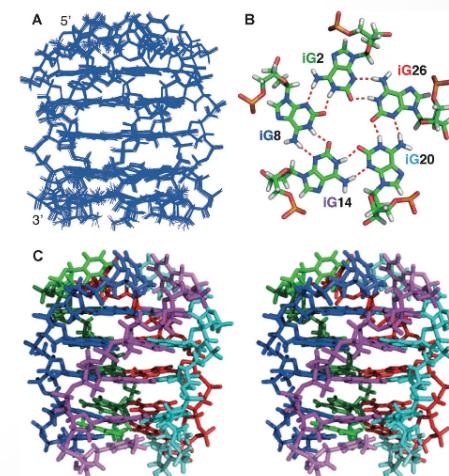


Non-canonical DNA structures

Name	Conformation	General Seq. Requirements	Sequence
Cruciform		Inverted Repeats	
Triplex		$(R \bullet Y)_n$ Mirror Repeats	
Slipped (Hairpin) Structure		Direct Repeats	
Tetraplex		Oligo (G) _n Tracts	$AG_3(T_2AG_3)_3$ single strand
Left-handed Z - DNA		$(YR \bullet YR)_n$	

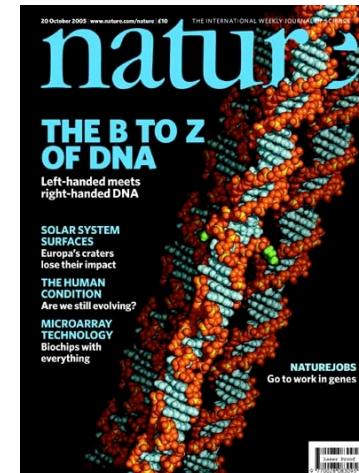
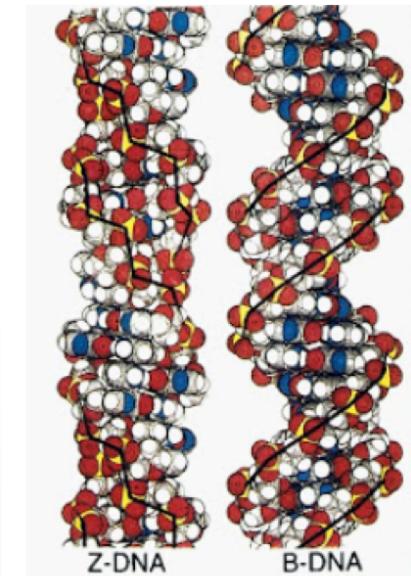


*Angewandte
Communications*

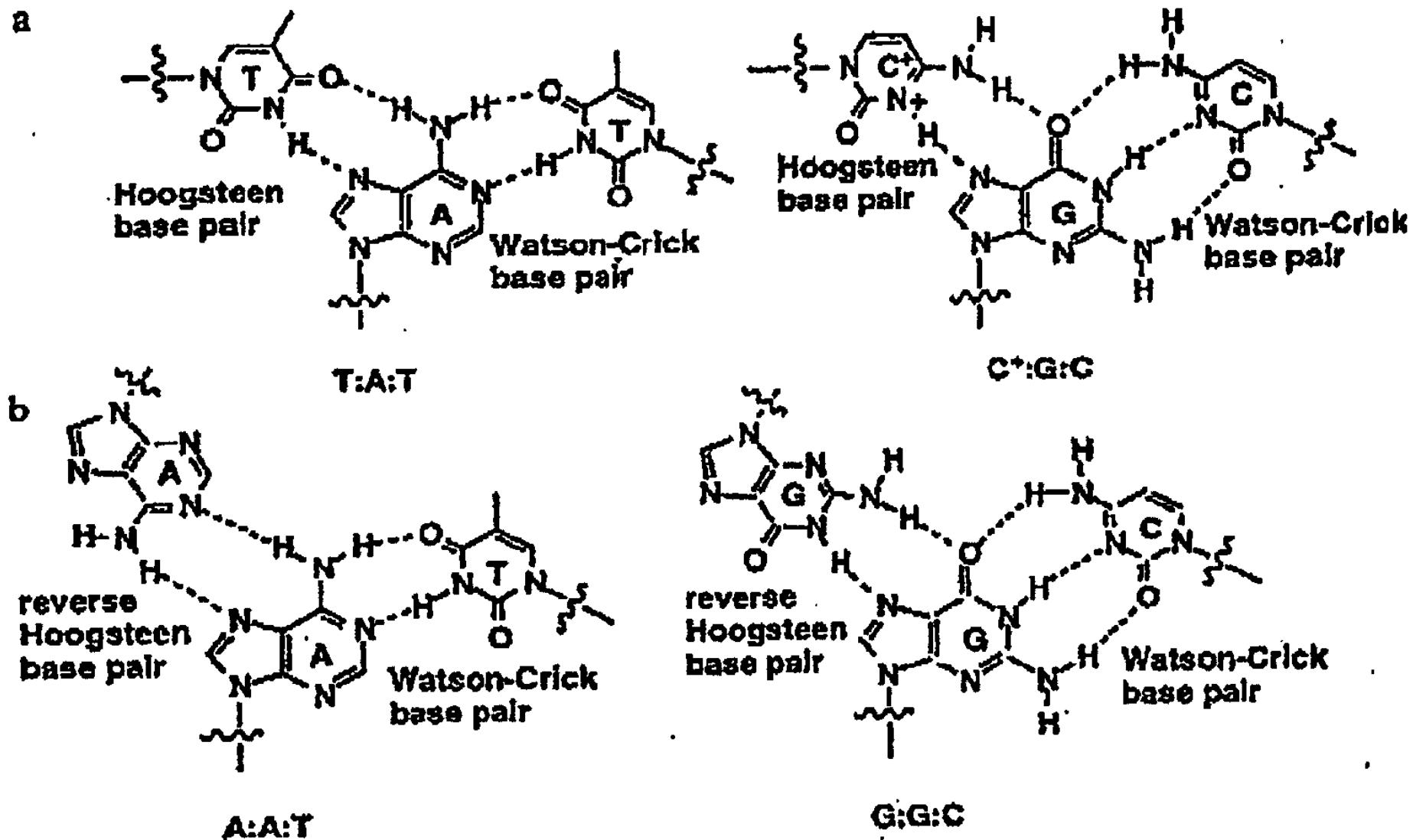


Z-DNA

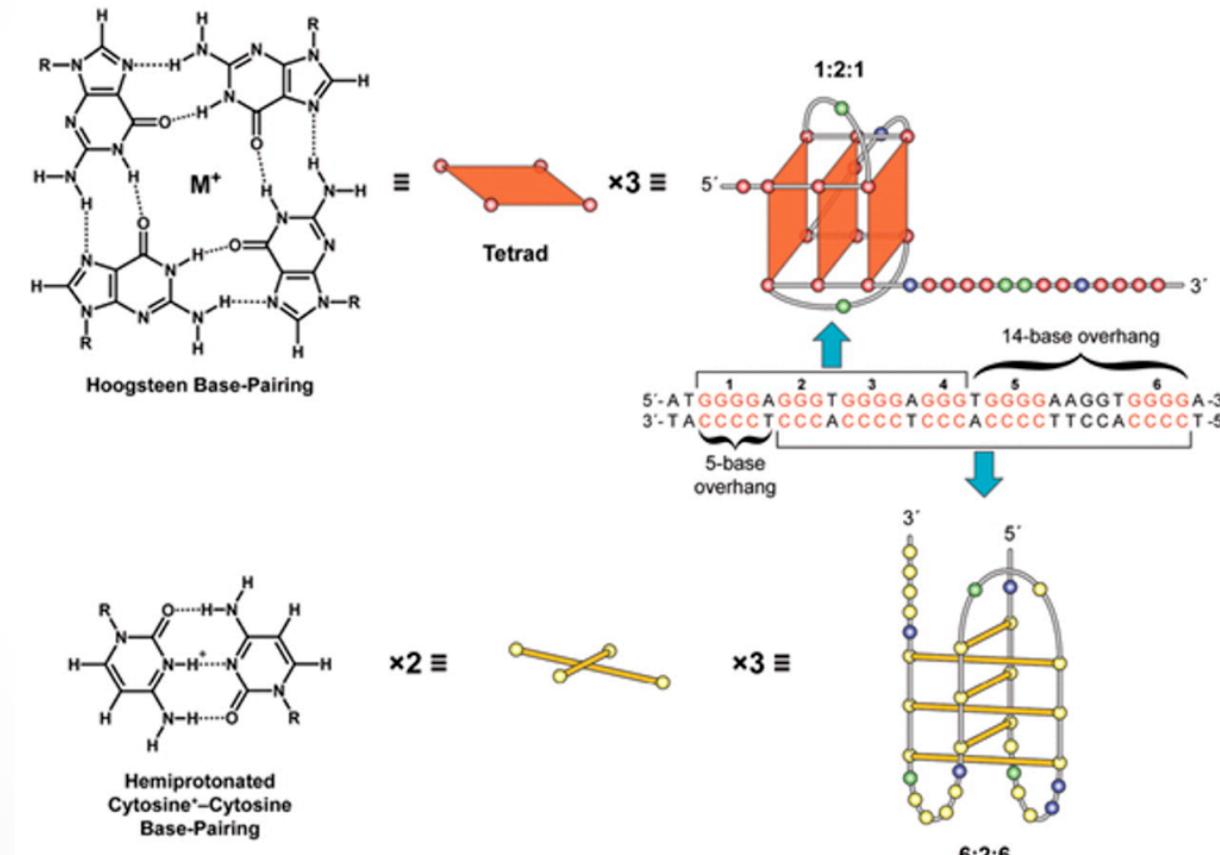
- Left-handed double helix
- T. Jovin first discovered it (CD).
- A. Rich first determined its structure (x-ray, Raman, etc)
- Z-DNA has Watson-Crick base pairs.



Both WC and HG basepairs form in triplex.



G-quadruplex, i-motif



Minireview: Non-B DNA Conformations

47413

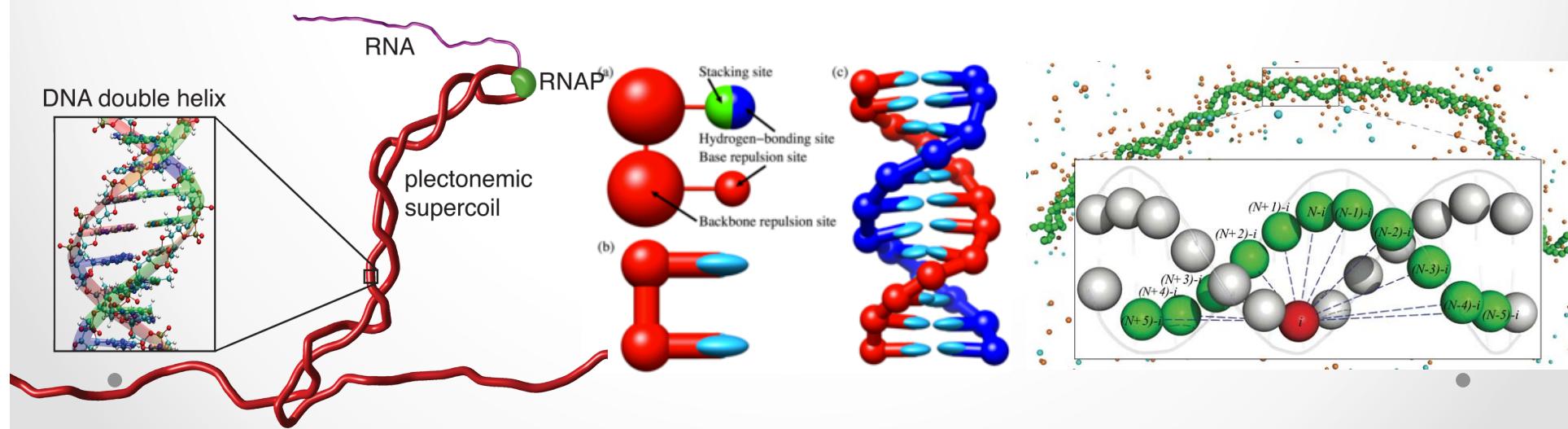
TABLE I
Selected genetic alterations, diseases, and DNA repeat motifs

DR with (R·Y), direct repeats composed of (R·Y)_n with mirror repeat symmetry.

Genetic alteration	Syndrome or metabolic event	DNA motifs	Ref.
t(11;22)(q23;q11.2)	Supernumerary der (22)	IR	17, 22, 23
t(17;22)(q11.2;q11.2)	Neurofibromatosis type 1	IR	17
t(1;22)(p21.2;q11.2)	Ependymoma	IR	17
t(4;22)(q35.1;q11.2)	Velocardiofacial	IR	17
i(17q)	Hematologic malignancies (chronic myeloid leukemia)	IR	19
del(Yq)	Spermatogenic failure	IR, DR	18, 26, 27
del(22)(q11.2q11.2)	DiGeorge, velocardiofacial, conotruncal anomaly face	IR, DR	20, 21
dup(22)(q11.2q11.2)	Cat-eye	IR, DR	20, 21
del(17)(p11.2p11.2)	Smith-Magenis	IR, DR	25, 28, 29
dup(17)(p11.2p11.2)	Smith-Magenis	IR, DR	25, 28, 29
dup(17)(p12p12)	Charcot-Marie-Tooth type 1	IR, DR	25, 28, 29
del(17)(p12p12)	Hereditary neuropathy with liability to pressure palsies	IR, DR	25, 28, 29
del(7)(q11.23q11.23)	Williams-Beuren	IR, DR	46
inv(7)(q11.23q11.23)	Predisposition to Williams-Beuren	IR	46
del(15)(q11q13)	Prader-Willi and Angelman	IR, DR	47
t(X;22)(q27;q11)	Myeloschizis and lumbosacral spina bifida	IR	48
Expanded (CTG-CAG) _n (coding)	Expanded polyglutamine diseases (Haw River, Huntington, Huntington's disease-like 2, spinobulbar muscular atrophy (Kennedy), spino-cerebellar ataxia (SCA) 1, SCA2, SCA3, SCA6, SCA7, SCA17)	DR	30, 31
Expanded (CTG-CAG) _n (non-coding)	Myotonic dystrophy type 1, SCA8, SCA12	DR	31
Expanded (CGG-CCG) _n (coding)	Expanded polyalanine diseases (infantile spasm, cleidocranial dysplasia, blepharophimosis/ptosis/epicanthus inversus type B, hand-foot-genital, synpolydactyly, oculopharyngeal muscular dystrophy, holoprosencephaly, oculopharyngeal muscular dystrophy)	DR	31, 49
Expanded (CGG-CCG) _n (non-coding)	Fragile XA, fragile XE, fragile XF, Jacobsen (FRA11B)	DR	30, 31
Expanded (GAA-TTC) _n	Friedreich's ataxia	DR with (R·Y)	30, 31
Expanded (GAC-GTC) _n	Pseudoachondroplasia, multiple epiphyseal dysplasia	DR	31
Expanded (CCTG-CAGG) _n	Myotonic dystrophy type 2	DR	31
Expanded (ATTCT-AGAAT) _n	SCA10	DR	31
Expanded (CCCCGCCCGCG) _n	Progressive myoclonus epilepsy type 1	DR	30, 31
Expanded 24-mer	Creutzfeldt-Jacob	DR	50
Contracted 3.3-kb D4Z4 repeat	Faciocapulohumeral muscular dystrophy	DR	51
Class switch recombination	IgH isotype switching	IR, G-rich	37-41

Atomic vs. coarse-grain pictures

- Atomic picture – PDB data
- Coarse-grain picture (The level of crudeness depends on what we are interested in and looking for.)
- More crude model for DNA – “uniform stiff” polymer
- Charged polymer surrounded by counter-ions



DNA mechanics

- DNA is highly charged → DNA is a stiff polymer
- Stiff unit segment contains many nucleotides and sequence information is averaged out.
- Sequence information does not matter
- On average, the same number of counter-ions are accumulated over DNA
-
-

DNA model: FJC

- Random walk
- A: 1D random walk, B: 3D random walk (with angle restriction and segment size a)
- cf. Freely-jointed chain model: links can rotate in arbitrary directions.

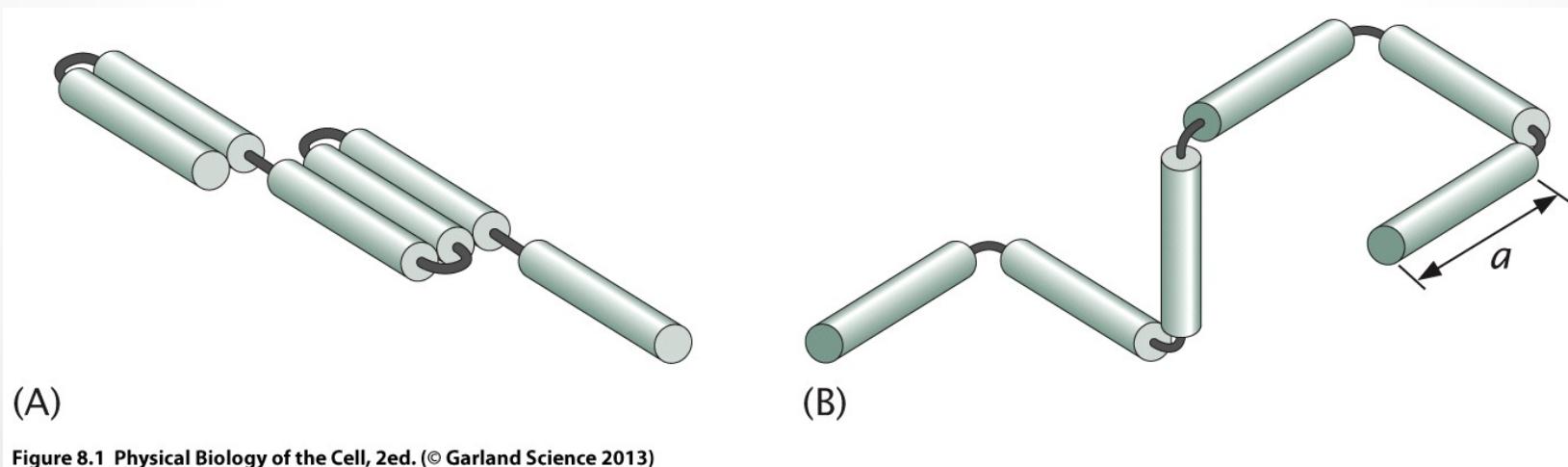


Figure 8.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Mean size of a random-walk macromolecules scales as \sqrt{N}

- In random walk models, every macromolecular configuration is equally probable.

$$\langle R \rangle = \left\langle \sum_{i=1}^N x_i \right\rangle = 0 \quad \langle x_i^2 \rangle = a^2$$

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

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

$$\langle R^2 \rangle = \langle \vec{R}_n^2 \rangle = \langle \vec{R}_n \cdot \vec{R}_n \rangle = \left\langle \left(\sum_{i=1}^n \vec{r}_i \right) \cdot \left(\sum_{j=1}^n \vec{r}_j \right) \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \langle \vec{r}_i \cdot \vec{r}_j \rangle$$

$$\vec{r}_i \cdot \vec{r}_j = l^2 \cos \theta_{ij}$$

$$\langle R^2 \rangle = l^2 \sum_{i=1}^n \sum_{j=1}^n \cos \theta^{|i-j|}$$

$$\langle R^2 \rangle = l^2 \sum_{i,j} \langle \hat{r}_i \cdot \hat{r}_j \rangle = l^2 \sum_{i=j} \langle \hat{r}_i \cdot \hat{r}_i \rangle = nl^2 \because \langle \hat{r}_i \cdot \hat{r}_j \rangle = 0 \text{ if } i \neq j$$

•

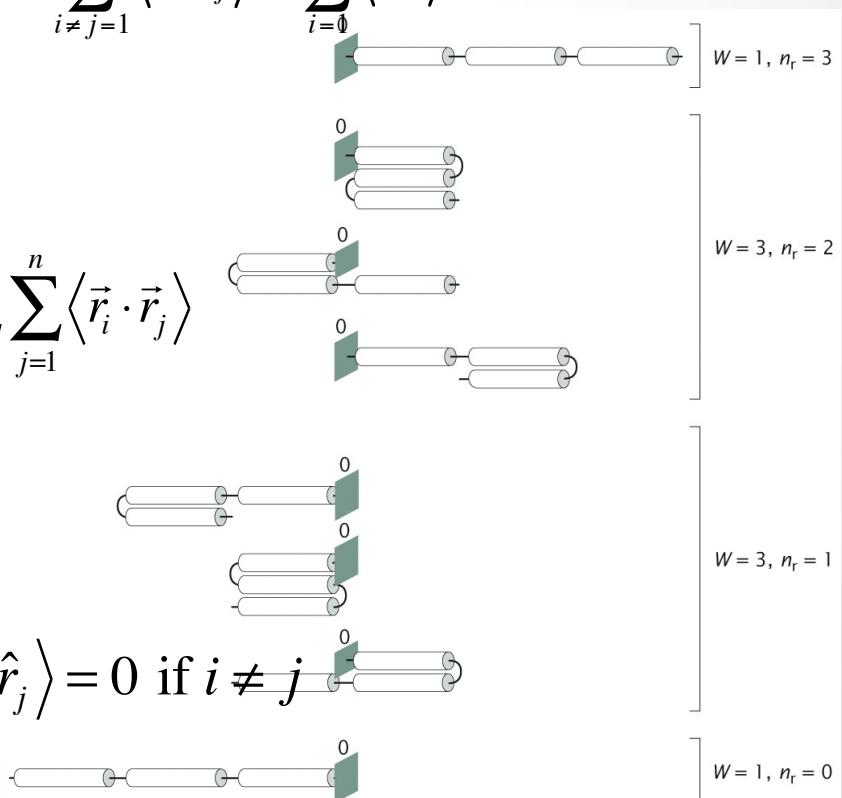


Figure 8.3 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

End-to-end probability distribution for a 1D “macromolecule”

- The probability of a given macromolecular configuration depends on its microscopic degeneracy.
- Calculation of the degeneracy based on probability theory.
- Question: what is the probability that n_r of N steps will be to the right (and n_l of N steps to the left) if $p_r = p_l = \frac{1}{2}$?
 - Probability of a particular sequence of N left and right steps: $(\frac{1}{2})^N$
 - Number of possibilities:

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

- Probability of taking n_r right steps:

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

-

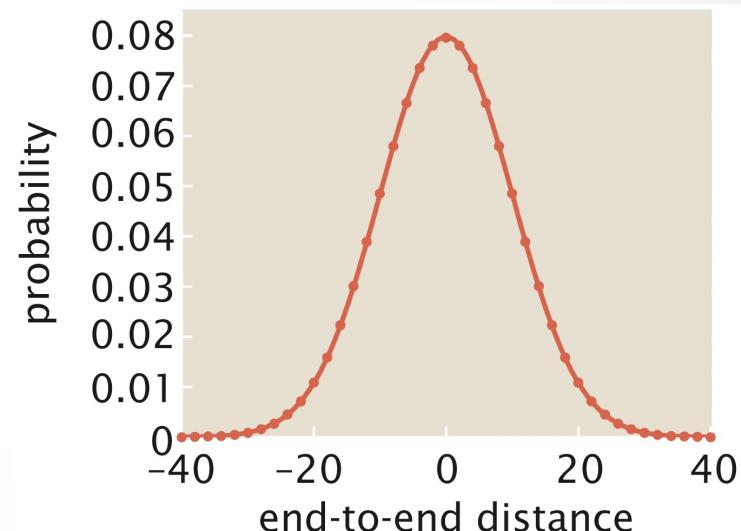


Figure 8.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

End-to-end probability distribution for a 1D “macromolecule”

- End-to-end distance: $R = (n_r - n_l) a$

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

$$\ln n! \approx n \ln n - n + \frac{1}{2} \ln(2\pi n)$$

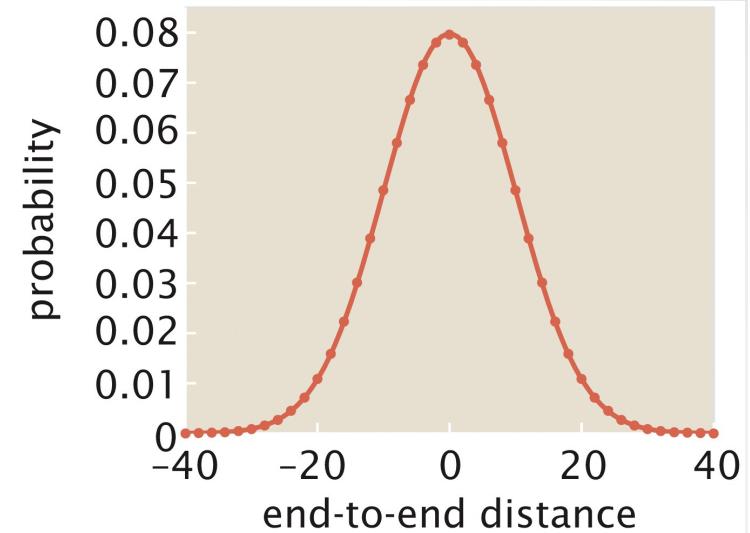
$$\ln p(R; N) = \dots = \ln 2 - \frac{1}{2} \ln(2\pi N) - \frac{R^2}{2Na^2}$$

$$p(R; N) = \frac{2}{\sqrt{2\pi N}} e^{-R^2/2Na^2} \quad \text{with } R \text{ being multiples of } 2a. \quad \leftarrow \text{Derive this.}$$

$$p(R; N) = \frac{1}{\sqrt{2\pi Na^2}} e^{-R^2/2Na^2} \quad \text{with } pdR \text{ being the probability}$$

that R falls within an interval of dR .

- $p(\mathbb{R}; N) = \left(\frac{3}{2\pi Na^2}\right)^{3/2} e^{-3R^2/2Na^2}$ for 3D random walk



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

- The persistence length is the scale over which the tangent-tangent correlation decays along the chain.
- It is the length scale over which “memory” of the tangent vector is lost in the presence of thermal fluctuation.

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

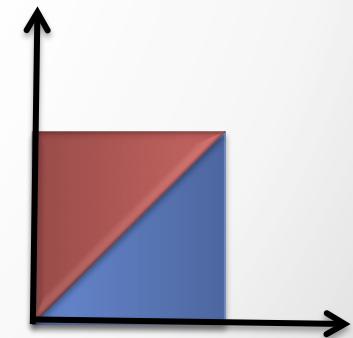
- In FJC model, Kuhn length a is twice the persistence length.

$$\mathbb{R} = \int_0^L ds \mathbf{t}(s)$$

$$\langle \mathbb{R}^2 \rangle = \left\langle \int_0^L ds \mathbf{t}(s) \int_0^L du \mathbf{t}(u) \right\rangle = 2 \int_0^L ds \int_s^L du e^{-(u-s)/\xi_p}$$

In the limit $L \gg \xi_p$, $u - s \rightarrow x$

- $\langle \mathbb{R}^2 \rangle \approx 2 \int_0^L ds \int_0^L dx e^{-x/\xi_p} = 2L\xi_p$



Random walk models for force-extension curves

$$\text{force} = -\frac{\partial G}{\partial L}$$

$$G(L) = -fL - k_B T \ln W(L; L_{tot}); L = (n_R - n_L)a; L_{tot} = Na$$

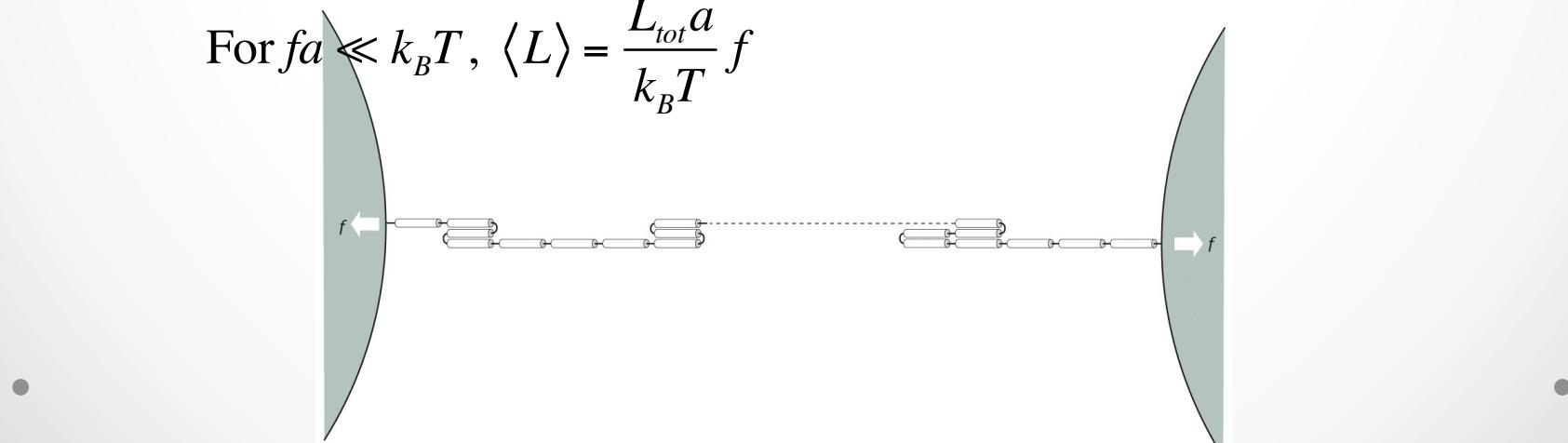
$$W(n_R; N) = \frac{N!}{n_R!(N - n_R)!}$$

$$G(n_R) = -2fn_Ra + k_B T(n_R \ln n_R + (N - n_R) \ln(N - n_R))$$

$$\frac{\partial G}{\partial n_R} = -2fa + k_B T \ln n_R - k_B T \ln(N - n_R) = 0$$

$$\frac{n_R}{n_L} = e^{2fa/k_B T} \rightarrow z = \frac{\langle L \rangle}{L_{tot}} = \frac{n_R - n_L}{n_R + n_L} = \tanh \frac{fa}{k_B T}$$

$$\text{For } fa \ll k_B T, \langle L \rangle = \frac{L_{tot}a}{k_B T} f$$



Force-extension curves by various FJC (freely-jointed chain) models

- A 3D FJC model (monomer points in one of six directions)

$$z = \frac{\langle L \rangle}{L_{tot}} = \frac{2 \sinh \beta fa}{4 + 2 \cosh \beta fa}$$

← Derive this.

- A 3D FJC model (off-lattice) $\langle L \rangle = Na \left(\coth \left(\frac{fa}{k_B T} \right) - \frac{k_B T}{fa} \right)$

← Derive this.

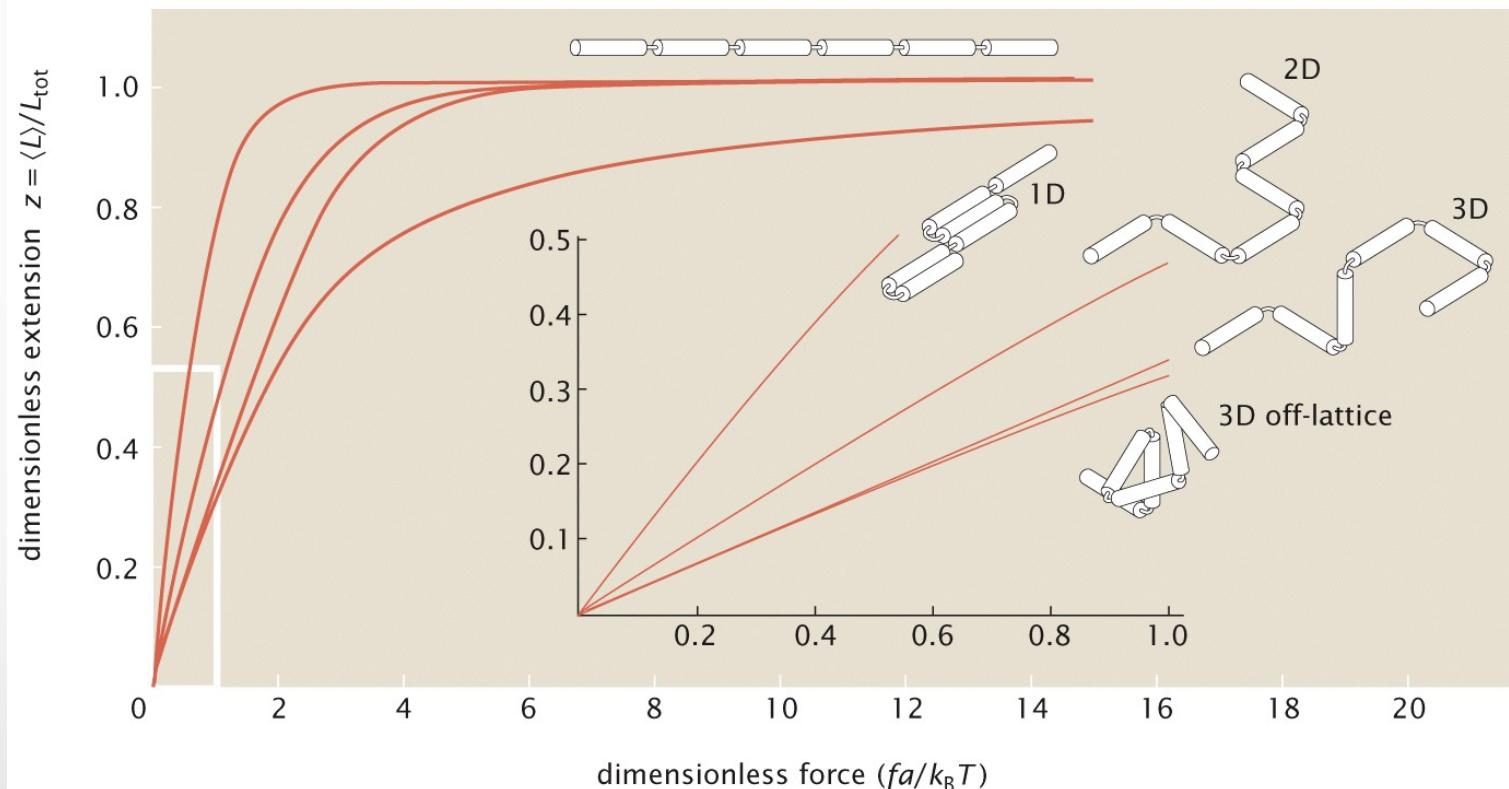


Figure 8.25 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

3D FJC under tension

$$Z = 1 + 1 + 1 + 1 + e^{-\beta fa} + e^{+\beta fa}$$

$$\langle z \rangle = \frac{\sinh \beta fa}{2 + \cosh \beta fa}$$

$$Z_1 = \int_0^{2\pi} d\phi \int_0^\pi e^{\beta fa \cos \theta} \sin \theta d\theta = 2\pi \int_{-1}^1 e^{\beta fa Q} dQ = 2\pi \frac{e^{\beta fa} - e^{-\beta fa}}{\beta fa} = 4\pi \frac{\sinh \beta fa}{\beta fa}$$

$$\langle z \rangle = \left[\coth \beta fa - \frac{1}{\beta fa} \right] \equiv \coth q - \frac{1}{q}$$

$$@ \beta fa \ll 1, \langle z \rangle = \left(\frac{1}{q} + \frac{q}{3} - \frac{q^3}{45} + \dots \right) - \frac{1}{q} \sim \frac{\beta fa}{3}; @ \beta fa \gg 1, \langle z \rangle \sim 1 - \frac{1}{q} \rightarrow \beta fa = \frac{1}{1 - \langle z \rangle}$$

Bending energy and persistence length

$$E_{bend} = \frac{K_{eff}}{2} \int_0^L \left| \frac{dt}{ds} \right|^2 ds$$

$$\hat{t} = (-\sin \theta, \cos \theta)$$

$$s = R\theta$$

$$\frac{d\hat{t}}{ds} = \frac{dt}{d\theta} \frac{d\theta}{ds} = (-\cos \theta, -\sin \theta) \frac{1}{R}$$

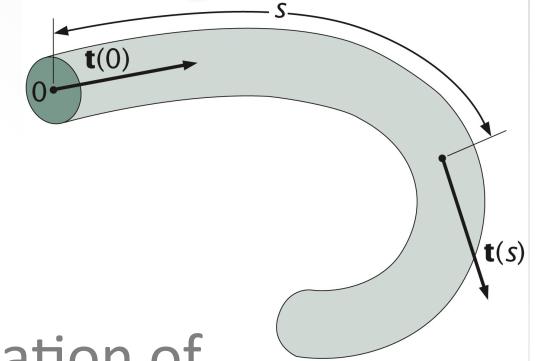


Fig. 10.1 Physical Biology of the Cell, 2nd Ed. (© Garland Science 2013)

- Thermal fluctuations tend to randomize the orientation of biological polymers.
 - Competition between thermal effects and deterministic forces:
Elastic forces set the length scale over which thermal fluctuations are tolerated.
- The persistence length is the length over which a polymer is roughly rigid.
 - Roughly speaking, ξ_p is the length of polymer whose radius of curvature is equal to its length.
- The persistence length characterizes the correlations in the tangent vectors at different positions along the polymer.
 - $$k_B T \approx \frac{K_{eff} L}{2R^2} \rightarrow \xi_p \approx \frac{K_{eff}}{2k_B T}$$
 - $$g(s) = \langle \mathbf{t}(\tau + s) \cdot \mathbf{t}(\tau) \rangle$$

$$g(s) = e^{-s/\xi_p} \approx 1 - \frac{s}{\xi_p}$$

$$E_{bend} = \frac{K_{eff}}{2s} \theta^2 \text{ where } \theta = \frac{s}{R}$$

$$g(s) = \langle \cos \theta(s) \rangle \approx \left\langle 1 - \frac{\theta^2(s)}{2} \right\rangle$$

$$\langle \theta^2(s) \rangle = \frac{1}{Z} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta \theta^2 e^{-\beta(K_{eff}/2s)\theta^2}$$

$$\text{where } Z = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta e^{-\beta(K_{eff}/2s)\theta^2}$$

$$\langle \theta^2(s) \rangle = -2s \frac{\partial \ln Z}{\partial (\beta K_{eff})}$$

Since $\sin \theta \approx \theta$ for small angles,

$$Z = \frac{2\pi s}{\beta K_{eff}} \int_0^\infty du e^{-u} = \frac{2\pi s}{\beta K_{eff}} \text{ where } u = \beta(K_{eff}/2s)\theta^2$$

Since $\partial \ln Z / \partial (\beta K_{eff}) = -1 / (\beta K_{eff})$,

$$g(s) = 1 - s / \beta K_{eff} \rightarrow \xi_p = \beta K_{eff}$$

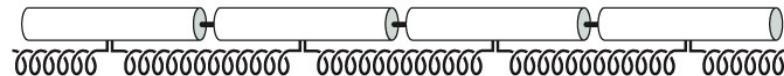
Persistence length vs. bending stiffness

- The persistence length is obtained by averaging over all configurations of the polymer.



Elasticity and entropy: The Worm-Like Chain

(A)



(B)

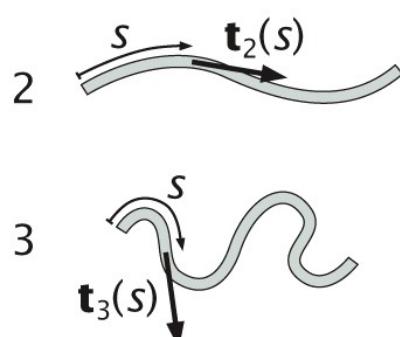


(C)

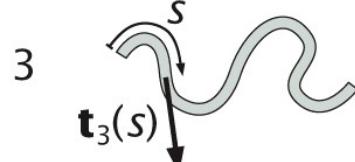
STATES



WEIGHTS



$$\exp\left[-\frac{\xi_p}{2} \int_0^L \left| \frac{d\mathbf{t}_2(s)}{ds} \right|^2 ds\right]$$



$$\exp\left[-\frac{\xi_p}{2} \int_0^L \left| \frac{d\mathbf{t}_3(s)}{ds} \right|^2 ds\right]$$

•
•
•

$$\langle z \rangle = \frac{1}{Z(f)} \int D\mathbf{t}(s) z \exp\left(-\frac{\xi_p}{2} \int_0^L \left| \frac{d\mathbf{t}(s)}{ds} \right|^2 ds + f \int_0^L t_z ds\right)$$

$$f\xi_p \approx \frac{z}{L} + \frac{1}{4(1 - z/L)^2} - \frac{1}{4}$$

← Derive this.



Figure 10.8 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

small force ($\beta fa \ll 1$)

$$\langle z^2 \rangle \equiv \left\langle \left(\hat{z} \cdot [\vec{R}(L) - \vec{R}(0)] \right)^2 \right\rangle = \frac{2\xi_p L}{3}$$

$$k \langle z^2 \rangle = \frac{f}{z} \langle z^2 \rangle = k_B T$$

$$\frac{f}{z} \frac{2\xi_p L}{3} = k_B T \rightarrow f = \frac{3k_B T}{2\xi_p L} z \rightarrow \frac{z}{L} = \frac{\beta fa}{3}$$

large force ($\beta f a \gg 1$)

$$\hat{t}(s) = \hat{z} t_{\parallel} + \vec{u}$$

$$t_{\parallel} = \sqrt{1 - |u|^2} = 1 - \frac{1}{2}|u|^2 + \dots$$

$$\beta E = -\beta f L + \int_0^L ds \left[\frac{\xi_p}{2} \left| \frac{du}{ds} \right|^2 + \frac{\beta f}{2} |u|^2 \right]$$

$$u_q = \int_0^L ds e^{iqs} u(s); u(s) = \frac{1}{L} \sum_q e^{-isq} u_q; \frac{du}{ds} = \frac{1}{L} \sum_q (-iq) e^{-isq} u_q$$

$$\beta E = -\beta fL + \frac{1}{2L} \sum_a (\xi_p q^2 + \beta f) |u_q|^2$$

$$\left\langle |u_q|^2 \right\rangle = \frac{2L}{(\xi_p q^2 + \beta f)} \text{ For a given } q, \text{ there are two transverse modes.}$$

$$\left\langle |u(s)|^2 \right\rangle = 2 \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{1}{(\xi_p q^2 + \beta f)} = \frac{1}{\sqrt{\beta \xi_p f}}$$

$$z = L \left\langle t_{||} \right\rangle = L \left(1 - \frac{1}{2} \left\langle |u|^2 \right\rangle + \dots \right) = L \left(1 - \frac{1}{\sqrt{4\beta \xi_p f}} + \dots \right)$$

For strong stretching, $\frac{f(2\xi_p)}{k_B T} \approx \frac{1}{2} \left(\frac{L}{L - \langle z \rangle} \right)^2 \leftarrow a = 2\xi_p$

At small relative extension, $2\beta f \xi_p \approx \frac{3\langle z \rangle}{L}$

Approximate interpolation: $\frac{f(2\xi_p)}{k_B T} \approx \frac{2\langle z \rangle}{L} + \frac{1}{2} \left(\frac{L}{L - \langle z \rangle} \right)^2 - \frac{1}{2}$

$$f = \frac{k_B T}{\xi_p} \left[\frac{z}{L} + \frac{1}{4} \left(\frac{1}{(1-z/L)^2} - 1 \right) \right]$$



Lecture 3

- Mechanical models: Worm Like Chain model
- DNA supercoils: definition (sign, magnitude)
- Linking number, twist, writhe
- Călugăreanu-White-Fuller theorem
- Energy associated with DNA supercoiling
- Non-canonical DNA structures (induced by SC)
-

Lecture 4

- Single molecule methods (revisit)
- Hybrid single molecule technique of smFRET & MT
- Case studies: DNA mechanics via single-molecule methods
- Case studies: Non-canonical DNA and its dynamics via single-molecule methods
-
-