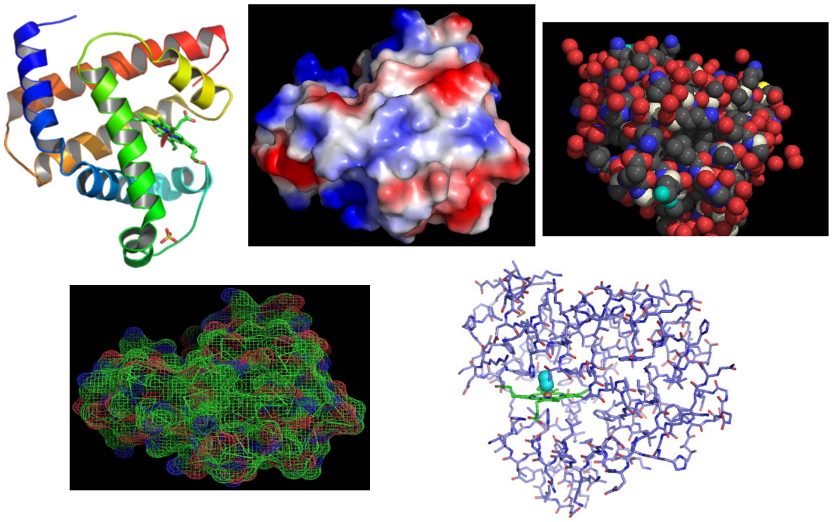
7. Statistical Description of Macromolecular Structure

There are a number of ways in which macromolecular structure is described in biophysics, which vary in type of information they are trying to convey. Consider these two perspectives on macromolecular structure that represent opposing limits: atomistic vs. statistical.

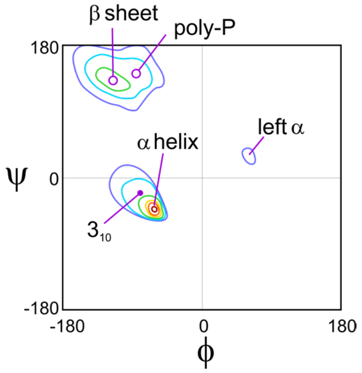
1) **Atomistic:** Use of atoms, small molecules, or functional groups as building blocks for biomolecular structure. This perspective is rooted in the dominant methods used for studying macromolecular structure (90% X-ray crystallography; 10% NMR). It has the most value for describing detailed Ångstrom to nanometer scale interactions of a chemical nature, but also tends to reinforce a unique and rigid view of structure, even though this cannot be the case at physiological temperatures.

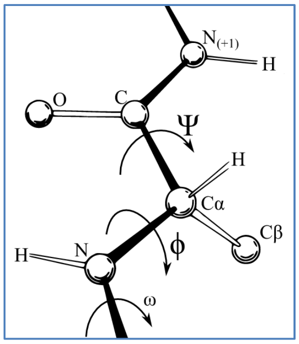
   The atomistic perspective is inherent to molecular force fields used in computational biophysics, which allow us to explore time-dependent processes and molecular disorder.

   Even within the atomistic representation, there are many complementary ways of representing macromolecular structure. Below are several representations of myoglobin structure, each is used to emphasize specific physical characteristics of the protein.



2) **Statistical/physical**: More applicable for disordered or flexible macromolecules. Emphasis is on a statistical description of molecules that can have multiple configurations. Often the atomic/molecular structure is completely left out. These tools have particular value for describing configurational entropy and excluded volume, and are influenced by the constraints of covalent bonding linkages along the chain. This approach is equally important: 30–40% of primary sequences in PDB are associated with disordered or unstructured regions. Conformational preferences are described statistically.

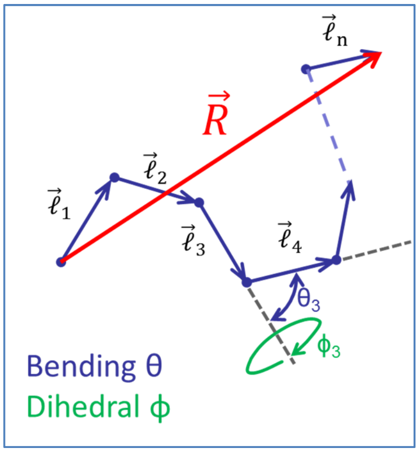




Statistical Models

- Structure described in terms of spatial probability distribution functions.
- There may be constraints on geometry or energy functions that describe interactions between and within chains.
- We will discuss several models that emerge for a continuous chain in space that varies in stiffness, constraints on conformation, and excluded volume.
    - Segment models: random coils, feely jointed chain, freely rotating chain
    - Lattice models: Flory–Huggins theory
    - Continuum model: worm-like chain

Segment Models[1]



- (n + 1) beads link by n segments or bonds of length ℓ.
- Each bead has a position $\vec{r}_i$ and a mass m_i.
- Each bond is assigned a vector, $\vec{\ell}_i = \vec{r}_i - \vec{r}_{i-1}$.
- The bending angle between adjacent segments i and (i + 1) is θ_i: $\cos\theta = \vec{\ell}_i \cdot \vec{\ell}_{i-1}$
- For each bending angle there is an associated dihedral angle ϕ_i defined as the rotation of segment (i+1) out of the plane defined by segments i and (i–1).
- There are n–1 separate bending and dihedral angles.

Statistical Variables for Macromolecules

End-to-end distance

The contour length is the full length of the polymer along the contour of the chain:

$$L_C = n\ell$$

Each chain has the same contour length, but varying dimensions in space that result from conformational flexibility. The primary structural variable for measuring this conformational variation is the end-to-end vector between the first and last bead, $\vec{R} = \vec{r}_n - \vec{r}_0$, or equivalently

$$\vec{R} = \sum_{i=1}^{n} \vec{\ell}_i$$

Statistically, the dimensions of a polymer can be characterized by the statistics of the end-to-end distance. Consider its mean-square value:

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

After expanding these sums, we can collect two sets of terms: (1) the self-terms with $i = j$ and (2) the interbond correlations ($i \neq j$):

---

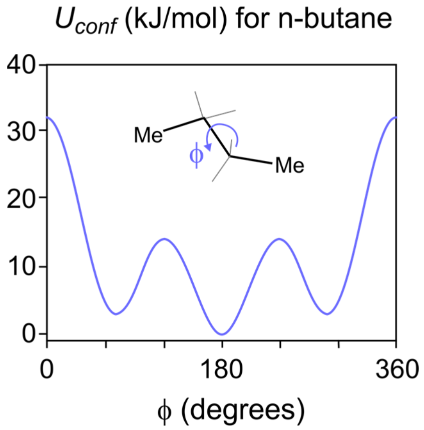
[1] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*. (W. H. Freeman, San Francisco, 1980), Ch. 18.; K. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. (Taylor & Francis Group, New York, 2010); P. J. Flory, *Principles of Polymer Chemistry*. (Cornell University Press, Ithaca, 1953).

$$\begin{aligned}\langle \vec{R}^2\rangle &= n\ell^2 + \sum_{j\neq i}\left\langle \vec{\ell}_i\cdot\vec{\ell}_j\right\rangle \\ &= n\ell^2 + \ell^2\sum_{j\neq i}\left\langle \cos\theta_{ij}\right\rangle\end{aligned} \tag{1}$$

Here θ_{ij} is the angle between segments i and j. This second term describes any possible conformational preferences between segments along the chain. We will call the factor $\langle\cos\theta_{ij}\rangle$ the segment orientation correlation function, which is also written

$$\begin{aligned} g(k) &= \langle\cos\theta_k\rangle \\ \theta_k &= \vec{\ell}_i\cdot\vec{\ell}_{i+k} \qquad\qquad k=|j-i| \end{aligned} \tag{2}$$

Here k refers to the separation between two segments. This correlation function can vary in value from 1 to –1, where +1 represents a highly aligned or extended chain and negative values would be very condensed or compact. No interband correlations ($g = 0$) is expected for placement of segments by a random walk.



Interbond correlation can be inserted into segment models, both through ad hoc rules, or by applying an energy function that constrains the intersegment interactions. For instance, the torsional energy function below, U_{conf}, would be used to weight the probability that adjacent segments adopt a particular torsional angle. A general torsional energy function $U_{\text{conf}}(\Theta)$ involves all $2(n-1)$ possible angles $\Theta = \{\theta_1,\phi_1,\theta_2,\phi_2,\ldots\theta_{n-1},\phi_{n-1}\}$, the joint probability density for adopting a particular conformation is

$$P(\Theta) = \frac{e^{-U_{\text{conf}}(\Theta)/k_BT}}{\int d\Theta\, e^{-U_{\text{conf}}(\Theta)/k_BT}}$$

The integral over Θ reflects $2(n-1)$ integrals over polar coordinates for all adjacent segments,

$$\int d\Theta = \int_0^{\pi}\int_0^{2\pi}\sin\theta_1 d\theta_1 d\phi_1 \cdots \int_0^{\pi}\int_0^{2\pi}\sin\theta_{n-1}d\theta_{n-1}d\phi_{n-1}$$

Then the alignment correlation function is

$$\left\langle \vec{\ell}_i\cdot\vec{\ell}_j\right\rangle = \ell^2\int d\Theta\,\cos\theta_{ij}\,P(\Theta)$$

This is not a practical form, so we will make simplifying assumptions about the form of this probability distribution. For instance, if any segments configuration depends only on its nearest neighbors then $P(\Theta)=P(\theta,\phi)^{(n-1)}$.

Persistence Length

For any polymer, alignment of any pair of vectors in the chain becomes uncorrelated over a long enough sequence of segments. To quantify this distance we define a "persistence length" ℓ_p.

$$\ell_p = \left\langle \hat{\ell}_i \cdot \sum_{j=1}^{n} \vec{\ell}_j \right\rangle \qquad \hat{\ell}_i = \frac{\vec{\ell}_i}{|\ell|}$$

This is the characteristic distance along the chain for the decay for the orientational correlation function between bond vectors,

$$g(k) = \ell^2 \langle \cos^k \theta \rangle$$

How will this behave? If you consider that $|\cos\theta| < 1$, then $\langle \cos^k\theta \rangle$ will drop with increasing k, approaching zero as $k \rightarrow \infty$. That is the memory of the alignment between two bond vectors drops with their separation, where the distance scale for the loss of correlation is ℓ_p. We thus expect a monotonically decaying form to this function:

$$g(k) = \ell^2 e^{-k\ell/\ell_p} \tag{3}$$

For continuous thin rod models of the polymer, this expression is written in terms of the contour distance s, the displacement along the contour of the chain (i.e., $s = \ell\, k$),

$$g(s) = \ell^2 e^{-|s|/\ell_p}$$

How do we relate θ and ℓ_p?[2] Writing $\langle cos^k\theta \rangle \approx \exp(k \ln[\langle \cos\theta \rangle])$ and equating this with eq. (3) indicates that

$$\ell_p = -\ell \ln\langle \cos \theta \rangle$$

For stiff chains, we can approximate $\ln(x) \approx (1-x)$, so

$$\ell_p \approx \frac{\ell}{1 - \langle \cos\theta \rangle}$$

Radius of gyration

The radius of gyration is another important structural variable that is closely related to experimental observables. Here the polymer dimensions are expressed as extension relative to the center of mass for the chain. This proves useful for branched polymers and heteropolymers (such as proteins). Denoting the position and mass of the i^{th} bead as $\vec{r}_i$ and m_i, we define the center of mass for the polymer as a mass-weighted mean position of the beads in space:

---

[2] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*. (W. H. Freeman, San Francisco, 1980), Ch. 19 p. 1033.

$$\vec{R}_0 = \frac{\sum_{i=0}^{n} m_i \vec{r}_i}{\sum_{i=0}^{n} m_i}$$

The sum index starting at *0* is meant to reflect the sum over *n+1* beads. The denominator of this expression is the total mass of the polymer $M = \sum_{i=0}^{n} m_i$. If all beads have the same mass, then $m_i/M = 1/(n+1)$ and R_0 is the geometrical mean of their positions.

$$\vec{R}_0 = \frac{1}{n+1}\sum_{i=0}^{n} \vec{r}_i$$

The radius of gyration R_G for a configuration of the polymer describes the mass-weighted distribution of beads R_0, and is defined through

$$\left\langle R_G^2 \right\rangle = \frac{1}{n+1}\sum_{i=0}^{n} \left\langle \vec{S}_i^2 \right\rangle$$

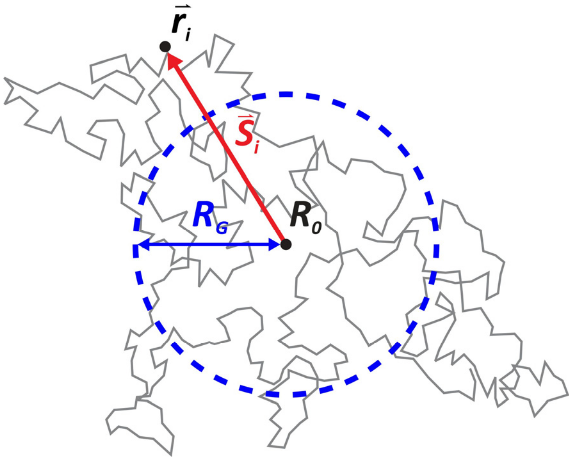
where $\vec{S}_i$ is gyration radius, i.e., the radial distance of the i^{th} bead from the center of mass

$$\vec{S}_i^2 = \frac{m_i}{M}\left(\vec{r}_i - \vec{R}_0\right)^2 \qquad \text{(mass-weighted)}$$

$$\vec{S}_i^2 = \frac{1}{n+1}\left(\vec{r}_i - \vec{R}_0\right)^2 \qquad \text{(equal mass beads)}$$

Additionally, we can show that the mean-squared radius of gyration is related to the average separation of all beads of the chain.

$$\left\langle R_G^2 \right\rangle = \frac{1}{(n+1)^2}\sum_{i=0}^{n}\sum_{j=0}^{n} \left\langle (\vec{r}_i - \vec{r}_j)^2 \right\rangle$$



Freely Jointed Chain

The freely jointed chain describes a macromolecule as a backbone for which all possible θ and ϕ are equally probable, and there are no correlations between segments. It is known as an "ideal chain" because there are no interactions between beads or excluded volume, and configuration of the polymer backbone follows a random walk. If we place the first bead at $r = 0$, we find that $\langle R\rangle = 0$, as expected for a random walk, and eq. (1) reduces to

$$\langle R^2\rangle = n\ell^2$$

or

$$R_{rms} = \langle R^2\rangle^{1/2} = \sqrt{n}\ell$$

While the average end-to-end distance may be zero, the variance in the end-to-end distribution is

$$\sigma_r = \sqrt{\langle R^2\rangle - \langle R\rangle^2} = \sqrt{n}\ell$$

The radius of gyration for an ideal chain is:

$$R_G = \sqrt{\frac{\langle R^2\rangle}{6}} = \sqrt{\frac{n\ell^2}{6}}$$

Gaussian Random Coil

The freely jointed chain is also known as a Gaussian random coil, because the statistics of its configuration are fully described by $\langle R\rangle$ and $\langle R^2\rangle$, the first two moments of a Gaussian end-to-end probability distribution $P(R)$.

The end-to-end probability density in one dimension can be obtained from a random walk with n equally sized steps of length ℓ in one dimension, where forward and reverse steps are equally probable. If the first bead it set at $x_0 = 0$, then the last bead is placed by the last step at position x. In the continuous limit:

$$P(x,n) = \sqrt{\frac{1}{2\pi n\ell^2}}\, e^{-x^2/2n\ell^2} \tag{4}$$

$P(x,n)\,dx$ is the probability of finding the end of the chain with n beads at a distance between x and $x+dx$ from its first bead. Note this equates the rms end-to-end distance with the standard deviation for this distribution: $\langle R^2\rangle = \sigma^2 = n\ell^2$.

To generalize eq. (4) to a three-dimensional chain, we recognize that propagation in the x, y, and z dimensions is equally probable, so that the 3D probability density can be obtained from a product of 1D probability densities $P(r) = P(x)P(y)P(z)$. Additionally, we need to consider the constraint that the distribution of end-to-end distances are equal in each dimension:

$$\langle \vec{R}^2\rangle = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = n\ell^2$$

and since $\sigma_x^2 = \sigma_y^2 = \sigma_z^2$,

$$\langle \vec{R}^2 \rangle = 3\sigma_x^2 = n\ell^2$$

Therefore,

$$P(r,n) = \sqrt{\frac{1}{2\pi\sigma_x^2}}\, e^{-x^2/2\sigma_x^2} \sqrt{\frac{1}{2\pi\sigma_y^2}}\, e^{-x^2/2\sigma_y^2} \sqrt{\frac{1}{2\pi\sigma_z^2}}\, e^{-x^2/2\sigma_z^2}$$

$$= \left(\frac{3}{2\pi\sigma^2}\right)^{3/2} e^{-3r^2/2\sigma^2}$$

To simplify, we define a scaling parameter with dimensions of inverse length

$$\beta = \sqrt{\frac{3}{2n\ell^2}} = \sqrt{\frac{3}{2}}\langle R^2 \rangle^{-1/2}$$

Then, the probability density in Cartesian coordinates,

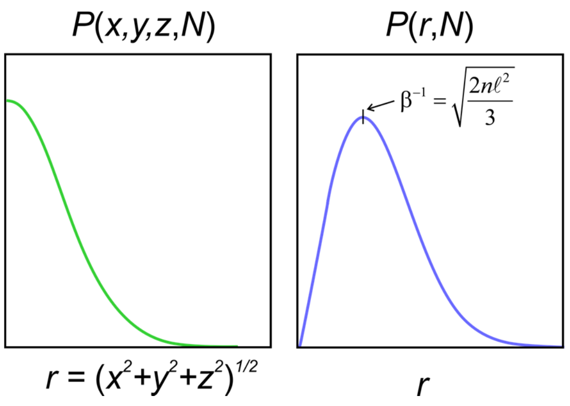
$$P(x,y,z,n) = \frac{\beta^3}{\pi^{3/2}} e^{-\beta^2 r^2} \qquad \text{where } r^2 = x^2 + y^2 + z^2$$

Note the units of $P(x,y,z,n)$ are inverse volume or concentration. The probability of finding the end of a chain of *n* beads in a box of volume *dx dy dz* at the position *x,y,z* is $P(x,y,z,n)\,dx\,dy\,dz$. This function illustrates that the most probable end-to-end distance for a random walk polymer is at the origin. On the other hand, we can also express this as a radial probability density that gives the probability of finding the end of a chain at a radius between *r* and *r*+*dr* from the origin. Since the volume of a spherical shell grows in proportion to its surface area:

$$P(r,n)dr = 4\pi r^2 P(x,y,z,n)dr$$

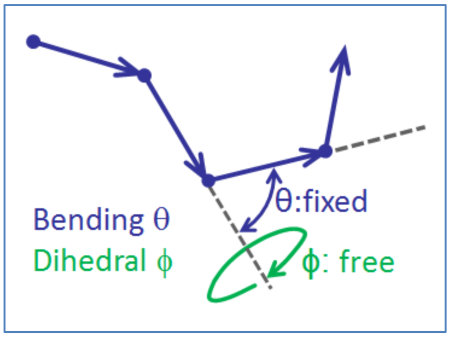
$$P(r,n) = 4\pi r^2 \left(\frac{3}{2\pi n\ell^2}\right)^{3/2} \exp\left[-\frac{3}{2}\frac{r^2}{n\ell^2}\right] \qquad (5)$$

The units of $P(r,n)$ are inverse length. For the freely jointed chain, we see that $\beta^{-1} = \sqrt{2\langle R^2 \rangle / 3}$ is the most probable end-to-end distance.



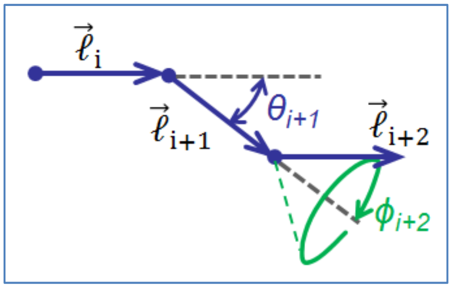
Freely Rotating Chain

An extension to the freely jointed chain that adds a single configurational constraint that better resembles real bonding in polymers is the freely rotating chain. In this case, the backbone angle θ has a fixed value, and the dihedral angle ϕ can rotate freely.



To describe the chain dimensions, we need to evaluate the angular bond correlations between segments. Focusing first on adjacent segments, we know that after averaging over all ϕ, the fixed θ assures that $\left\langle \vec{\ell}_i \cdot \vec{\ell}_{i+1} \right\rangle = \ell^2 \cos\theta$. For the next segment in the series, only the component parallel to $\vec{\ell}_{i+1}$ will contribute to sequential bond correlations as we average over ϕ_{i+2}:

$$\begin{aligned}\left\langle \vec{\ell}_i \cdot \vec{\ell}_{i+2} \right\rangle &= \left\langle \cos(\theta_i)\cos(\theta_{i+1}) - \sin(\theta_i)\sin(\theta_{i+1})\cos(\phi_{i+1}) \right\rangle \\ &= \ell^2 \cos^2\theta\end{aligned}$$



Extending this reasoning leads to the observation

$$\left\langle \vec{\ell}_i \cdot \vec{\ell}_j \right\rangle = \ell^2 (\cos\theta)^{j-i}$$

To evaluate the bond correlations in this expression, it is helpful to define an index for the separation between two bond vectors:

$$k = j - i$$

and

$$\alpha = \cos\theta$$

Then the segment orientation correlation function is

$$g(k) = \left\langle \vec{\ell}_i \cdot \vec{\ell}_j \right\rangle = \ell^2 \alpha^k$$

For a separation k on a chain of length n, there are n–k possible combinations of bond angles,

$$\sum_{j\neq i}\left\langle (\cos\theta)^{j-i}\right\rangle = \sum_{k=1}^{n-1}(n-k)\alpha^k$$

$$\therefore \qquad \langle R^2\rangle = n\ell^2 + \ell^2\sum_{k=1}^{n-1}(n-k)\alpha^k$$

From this you can obtain

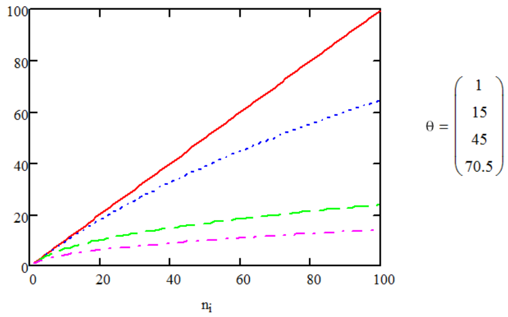
$$\langle R^2\rangle = n\ell^2\left(\frac{1+\alpha}{1-\alpha} - \frac{2\alpha\left(1-\alpha^n\right)}{n(1-\alpha)^2}\right)$$

In the limit of long chains (n → ∞), we find

$$\langle R^2\rangle \to n\ell^2\left(\frac{1+\alpha}{1-\alpha}\right)$$

and

$$R_G = \sqrt{\frac{n\ell^2}{6}\left(\frac{1+\alpha}{1-\alpha}\right)}$$



RMS end-to-end distance $\langle R^2\rangle^{1/2}$ in units of ℓ as a function of n and θ

Restricted dihedrals

When the freely rotating chain is also amended to restrict the dihedral angle ϕ, we can solve the mean square end-to-end distance in the limit $n \to \infty$. Given an average dihedral angle,

$$\beta = \langle\cos\phi\rangle$$

$$\langle R^2\rangle = n\ell^2\left(\frac{1+\alpha}{1-\alpha}\right)\left(\frac{1+\beta}{1-\beta}\right)$$

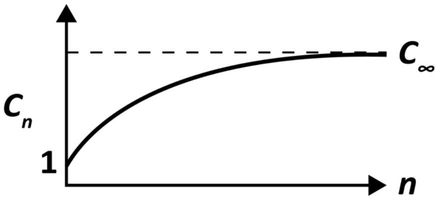
Nonideal Behavior

Flory characteristic ratio

Real polymers are stiff and have excluded volume, but the $R \sim \sqrt{n}$ scaling behavior usually holds at large n ($R \gg \ell_p$). To characterize non-ideality, we use the Flory characteristic ratio:

$$C_n = \frac{\langle R^2 \rangle}{n\ell^2}$$

For freely jointed chains $C_n = 1$. For nonideal chains with angular correlations, $C_n > 1$. C_n depends on the chain length n, but should have an asymptotic value for large n: C_∞. For example, if we examine long freely rotating chains



$$C_\infty = \lim_{n\to\infty} \frac{\langle R^2 \rangle}{n\ell^2} = \frac{1+\alpha}{1-\alpha} \qquad \alpha = \cos\theta$$

(In practice, this limit typically holds for $n > 30$). Consider a tetrahedrally bonded polymer with full angle 109° ($\theta = 54°$). then $\cos\theta = 1/3$, and $C_n = 2$. In practice, we reach the long chain limit C_∞ at $n \approx 10$. This relation works well for polyglycine and polyethylene glycol (PEG).

Statistical segment or Kuhn length

How stiff or flexible a polymer is depends on the length scale of observation. What is stiff on one scale is flexible for another. For an infinitely long polymer, one can always find a length scale for which its statistics are that of a Gaussian random coil. As a result for a segment polymer, one can imagine rescale continuous segments into one longer "effective segment" that may not represent atomic dimensions, but rather is defined in order to correspond to a random walk polymer, with $C_n = 1$. Then, the effective length of the segment is ℓ_e (also known as the Kuhn length) and the number of effective segments is n_e. Then the freely jointed chain equations apply:

$$L_C = n_e \ell_e$$

$$\langle R^2 \rangle = n_e \ell_e^2$$

From these equations, $\ell_e = \langle R^2 \rangle / L_C$. We see that $\ell_e \gg \ell$ applies to stiff chains, whereas $\ell_e \approx \ell$ are flexible.

We can also write the contour length as $L_C = \gamma n \ell$, where γ is a geometric factor < 1 that describes constraint on bond angles. For a freely rotating chain: $\gamma = \cos(\theta/2)$. Using the long chain chain expressions ($n\to\infty$): $\langle R^2 \rangle = C_\infty n \ell^2$, we find

$$\ell_e = \left(\frac{C_\infty}{\gamma}\right)\ell$$

$$n_\ell = \left(\frac{\gamma^2}{C_\infty}\right)n$$

$$\ell_p = \left(\frac{C_\infty + 1}{2}\right)\ell$$

Representative values for polymer segment models

	C_∞	(n_e/n)	ℓ (nm)	ℓ_e (nm)	γ	ℓ_p(nm)
Polyethylene	6.7	(n > 10)	0.154	1.24	0.83	
PEG	3.8			0.34		
Poly·alanine	9	(n > 70)	0.38	3.6	0.95	0.5
Poly·proline	90	(n > 700)				5-10
dsDNA	86		0.35	30–100	1	50
ssDNA						1.5
Cellulose						6.2
Actin				16700		10000–20000

Excluded Volume Effects

In real polymers, the chance of colliding with another part of the chain increases with chain length.

$$\langle R^2 \rangle = n\ell^2 + \sum_{i \neq j} \langle \vec{\ell}_i \cdot \vec{\ell}_j \rangle$$

$$\langle \vec{\ell}_i \cdot \vec{\ell}_j \rangle = g(s) = \langle \vec{\ell}_i \cdot \vec{\ell}_{i+s} \rangle \qquad s = |i-j|$$

$g(s)$ gives the orientational correlations between polymer segments.

Flory, statistical mechanics of chain molecules

- If correlations are purely based on bond angles and rotational potential, then $g(s)$ decays exponentially with s. There is no excluded volume.
- With excluded volume, $g(s)$ does not vanish for large k. There are "long-range" interactions within the chain.
    - "Long range" means along long distance along contour, but short range in space.
- Excluded volume depends on chain + solvent and temperature.

Virial expansion

At low densities, thermodynamic functions can be expanded in a power series in the number of particles per unit volume: $n = N/V$ (density).

$$F = F^0 + F_{\text{int}}$$

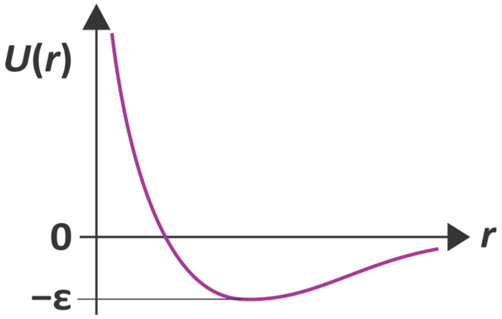
$$F_{\text{int}} = N_p k_B T\left(nB + n^2C + \ldots\right)$$

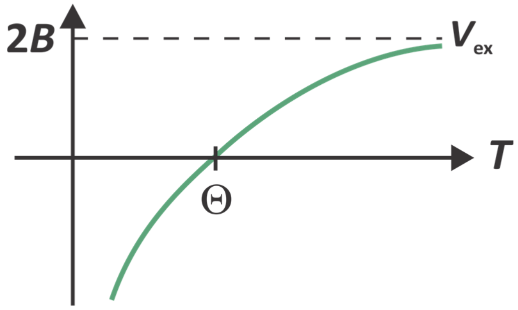
- F^0 refers to ideal chain
- N_p is # of polymer molecules
- B: units of volume

Excluded volume (repulsion) and attractive interactions are related to the second virial coefficient B. The excluded volume (or volume correlation relative to ideal behavior) for interacting beads of a polymer chain is calculated from

$$V_{\text{ex}} = \int d^3r \left(1 - \exp\left[-U(r)/k_BT\right]\right)$$

$U(r)$ is the interaction potential. In the high temperature limit $V_{\text{ex}} = 2B$. So $2B$ can be associated with the excluded volume associated with one segment (bead) of the chain.





Temperature dependence

- At high T $\left(k_B T \gg \varepsilon\right)$
  The attractive part of potential is negligible, and repulsions result in excluded volume. In this limit $2B \approx V_{ex}$.
- As $T \to 0$, the attractive part of potential matters more and more, resulting in collapse relative to ideal chain.
- Cross over: Theta point T=Θ

  Near Θ $\qquad 2B \sim V_{ex}\left(\frac{T-\Theta}{\Theta}\right)$

  $T > \Theta$ High T. Repulsion dominates. Polymer swells (good solvent)

  $T < \Theta$ Low T. Attractions dominate. Polymer collapses (globule, poor solvent)

Polymer swelling

At high temperatures (T≫Θ), the free energy of a coil can be expressed in terms interaction potential, which is dominated by repulsions that expand the chain, and the entropic elasticity that opposes it (see next chapter).

$$F = U - TS = nk_B T\, B\frac{3n}{4\pi R^3} + k_B T\frac{3R^2}{2n\ell^2} + const.$$

By minimizing F with respect to the end-to-end distance, R, and solving for R, we can find how the R scales with polymer size:

$$R \propto (B\ell^2)^{3/5} n^{3/5}$$

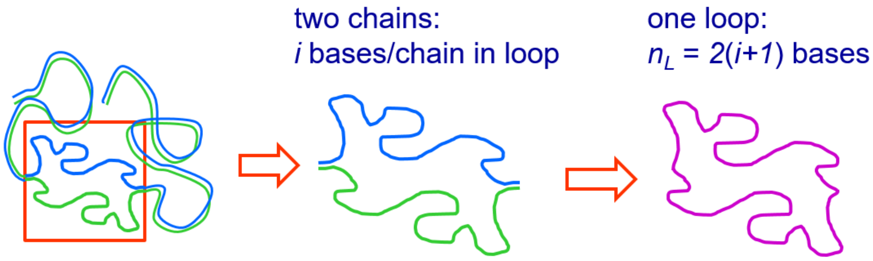
We see that the end-to-end distance of the chain with excluded volume scales with monomer number (n) with a slightly larger exponent than an ideal chain: $n^{3/5}$ rather than $n^{1/2}$. Generally, the relationship between R and n is expressed in terms of the Flory exponent, ν, which is related to several physical properties of polymer chains:

$$R \propto n^{\nu}$$

Polymer Loops

For certain problems, we are concerned with cyclic polymers chains:

- Bubbles/loops in DNA melting
- Polypeptide and RNA hairpins
- DNA strand separation in transcription
- Cyclic DNA, chromosomal looping, and supercoiling



In describing macromolecules in closed loop form, the primary new variable that we need to address is the loop's configurational entropy. Because of configurational constraints that tie the ends of a loop together ($R_{ee} \rightarrow 0$) the loop has lower configurational entropy than an unrestrained coil.

Let's describe how the configurational entropy of a loop S_L depends on the size of the loop. We will consider the segment model with n_L segments in the loop. We start with the radial probability distribution for an unconstrained random coil, which is the reference state for our calculations:

$$P(r,n) = 4\pi r^2 \left(\frac{3}{2\pi n\ell^2}\right)^{3/2} \exp\left[-\frac{3}{2}\frac{r^2}{n\ell^2}\right] \tag{6}$$

The entropy of the loop S_L will reflect the constraints placed by holding the ends of the random coil together, which we describe by saying the ends of the chain must lie within a small distance Δr of each other. Since $R_{ee} < \Delta r$, $\Delta r^2 \ll n\ell^2$, and the exponential term in eq. (6) is ~*1*. Then the probability of finding a random coil configuration with an end-to-end distance within a radius Δr is

$$\begin{aligned} P_L(n_L) &\approx \int_0^{\Delta r} dr\, 4\pi r^2 \left(\frac{3}{2\pi n_L \ell^2}\right)^{3/2} \\ &= \left(\frac{6}{\pi}\right)^{1/2} \left(\frac{\Delta r}{\ell}\right)^3 n_L^{-3/2} \\ &\equiv b n_L^{-3/2} \end{aligned}$$

In the last line we find that the probability of finding a looped chain decreases as $P_L \propto n_L^{-3/2}$, where b is the proportionality constant that emerges from integration. From the assumptions we made, $b \ll 1$, and $P_L < 1$.

To calculate the configurational entropy of the chain, we assume that the polymer (free or looped) can be quantified by Ω configurational states per segment of the chain. This reflects the fact that our segment model coarse-grains over many internal degrees of freedom of the macromolecule. Then, the entropy of a random coil of n segments is $S_C = k_B \ln \Omega^n$. To calculate the loop entropy, we correct the unrestrained chain entropy to reflect the constraints placed by holding the ends of the random coil together in the loop.

$$S_L = S_C + k_B \ln P_L$$

This expression reflects the fact that the number of configurations available to the constrained chain is taken to be $\Omega_L(n_L) = \Omega^{n_L} P_L\left(n_L\right)$, and each of these configurations are assumed to be equally probable ($S_L = k_B \ln \Omega_L$). Since $P_L < 1$, the second term is negative, lowering the loop entropy relative to the coil. We find that we can express the loop configurational entropy as

$$S_L(n_L) = k_B \left[n_L \ln \Omega - b - \frac{3}{2} \ln n_L \right]$$

Since this expression derives from the random coil, it does not account for excluded volume of the chain. However, regardless of the model used to obtain the loop entropy, we find that we can express it is the same form:

$$S_L(n_L) = k_B \left[n_L a - b + c \ln n_L \right]$$

where a, b, and c are constants. For the random coil $c = 1.50$, and for a self-avoiding random walk on a cubic lattice we find that it increases to $c = 1.77$. In 2D, a random coil results in $c = 1.0$, and a SAW gives $c = 1.44$.

Readings

1. M. Rubinstein and R. H. Colby, *Polymer Physics*. (Oxford University Press, New York, 2003).
2. K. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. (Taylor & Francis Group, New York, 2010).
3. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*. (W. H. Freeman, San Francisco, 1980).
4. R. Phillips, J. Kondev, J. Theriot and H. Garcia, *Physical Biology of the Cell*, 2nd ed. (Taylor & Francis Group, New York, 2012).
5. P. J. Flory, *Principles of Polymer Chemistry*. (Cornell University Press, Ithaca, 1953).