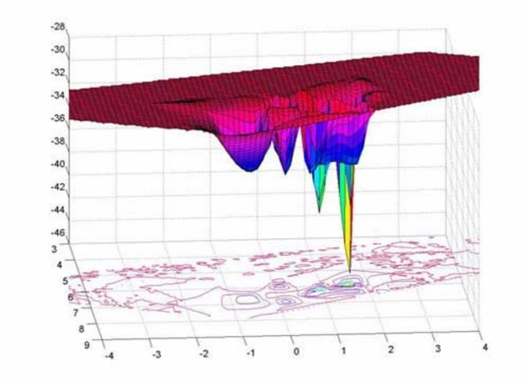
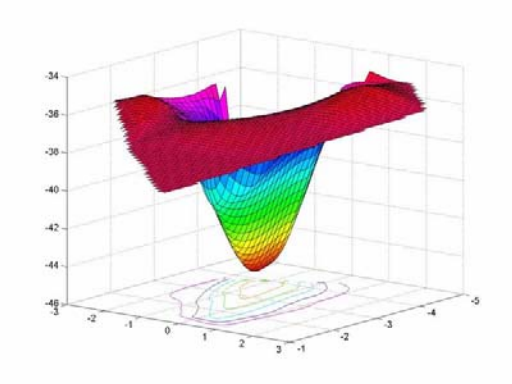
Protein dynamics



B. Ozkan, K.A. Dill & I. Bahar, Protein Sci. 11, 1958-1970, 2002

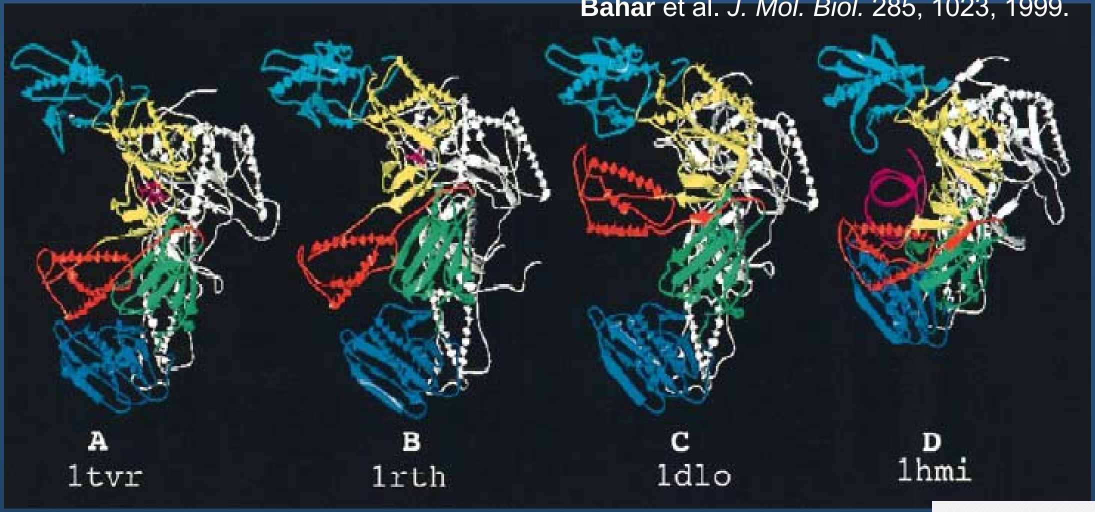
- Folding/unfolding dynamics
  - Passage over one or more energy barriers
  - Transitions between infinitely many conformations
- Fluctuations near the folded state
  - Local conformational changes
  - Fluctuations near a global minimum



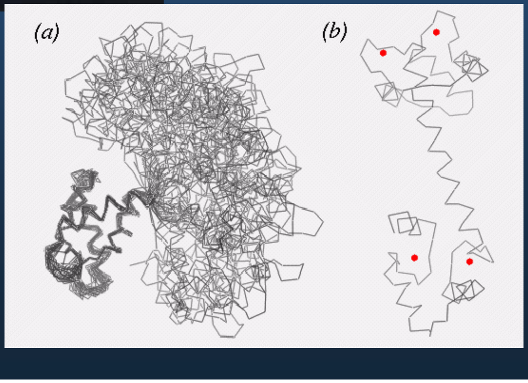
Stuctures suggest mechanisms of function

A. Comparison of static structures available in the PDB for the same protein in different form has been widely used as an ***indirect*** method of inferring dynamics.

Bahar et al. *J. Mol. Biol.* 285, 1023, 1999.



B. NMR structures provide information on fluctuation dynamics



Several modes of motions in native state

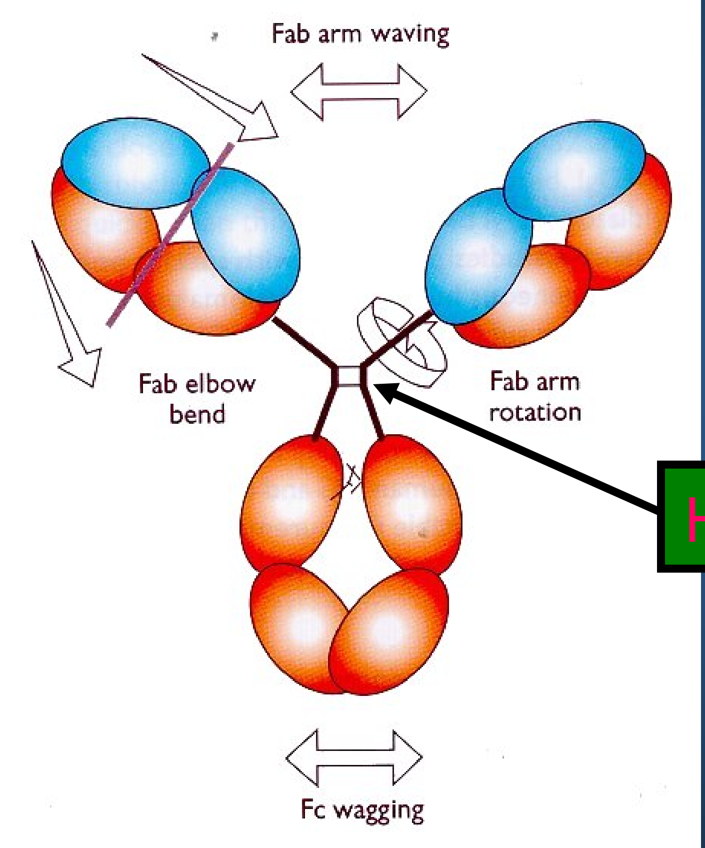
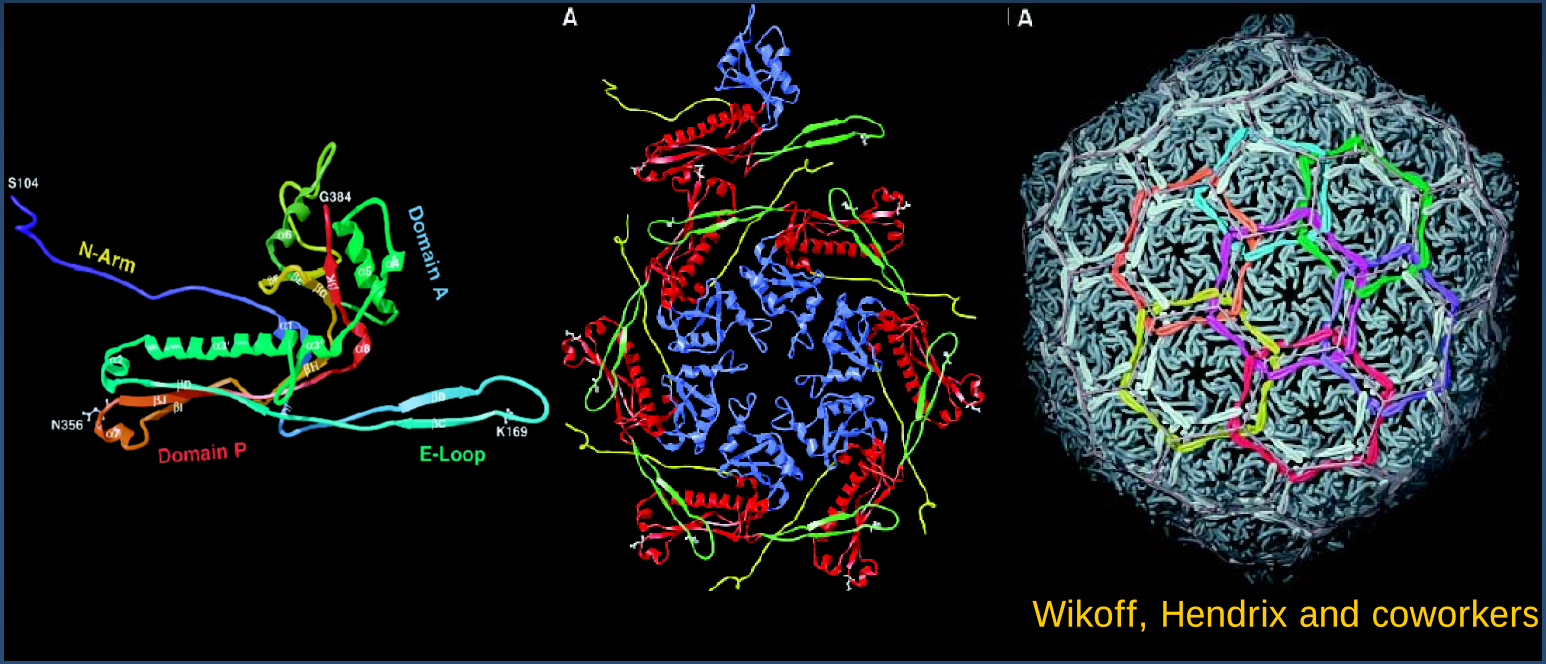


Figure 4 *Flexibility of the IgG molecule.*
Reproduced with permission from Immunology Today, February 1995[22].

Hinge site

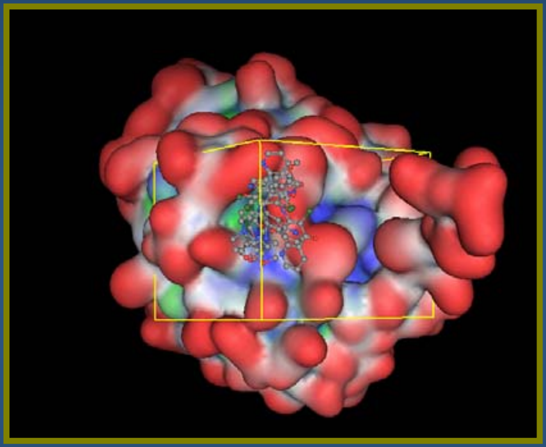
Supramolecular dynamics



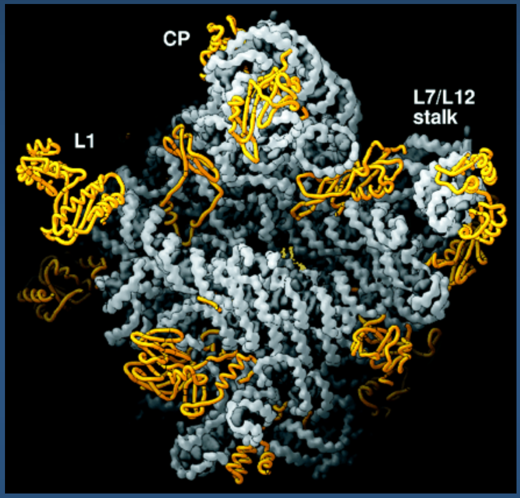
Wikoff, Hendrix and coworkers

Multiscale modeling – from full atomic to multimeric structures

Progresses in molecular approaches:
Coarse-grained approaches for large complexes/assemblies

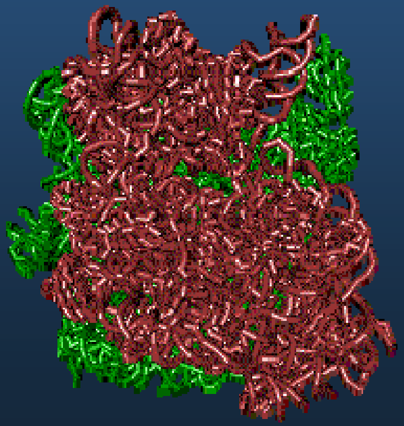


← -------- **25 Å** ----- →

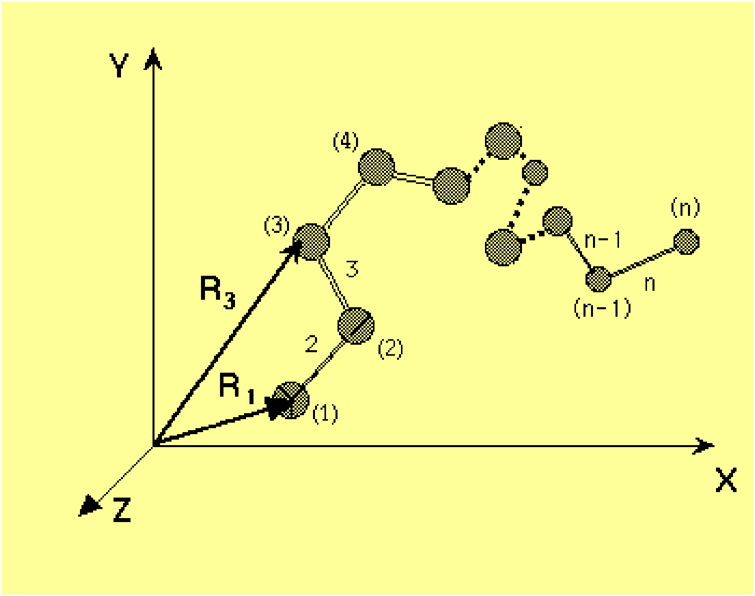


← -------- **250 Å** ------ →

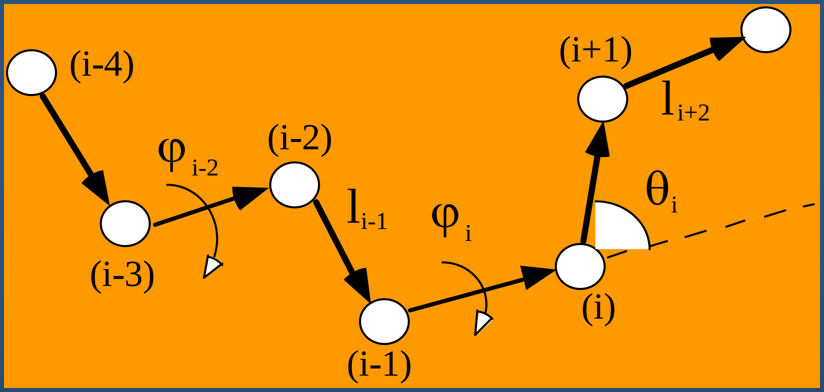
Example: EN models for modeling ribosomal machinery (Frank et al, 2003; Rader et al., 2004)



Macromolecular Conformations



Schematic representation of a chain of n backbone units. Bonds are labeled from 2 to n, and structural units from 1 to n. The location of the ith unit with respect to the laboratory-fixed frame OXYZ is indicated by the position vector **R**i.



Schematic representation of a portion of the main chain of a macromolecule. **l**i is the bond vector extending from unit i-1 to i, as shown. φ_i denotes the torsional angle about bond i.

How/why does a molecule move?

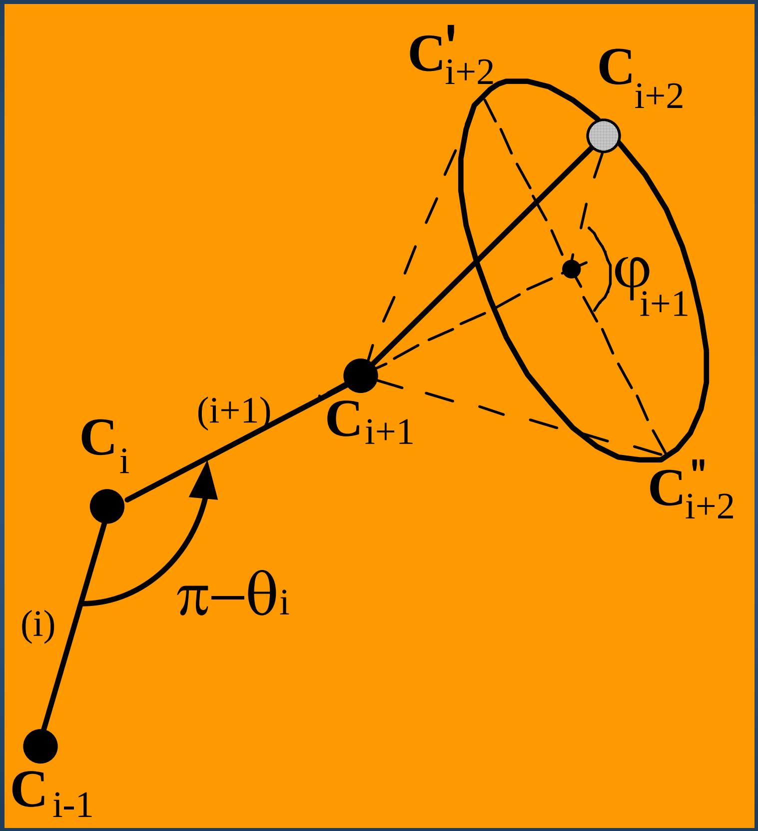
Among the 3N-6 internal degrees of freedom, **bond rotations** (i.e. changes in dihedral angles) are the softest, and mainly responsible for the functional motions

Two types of bond rotational motions

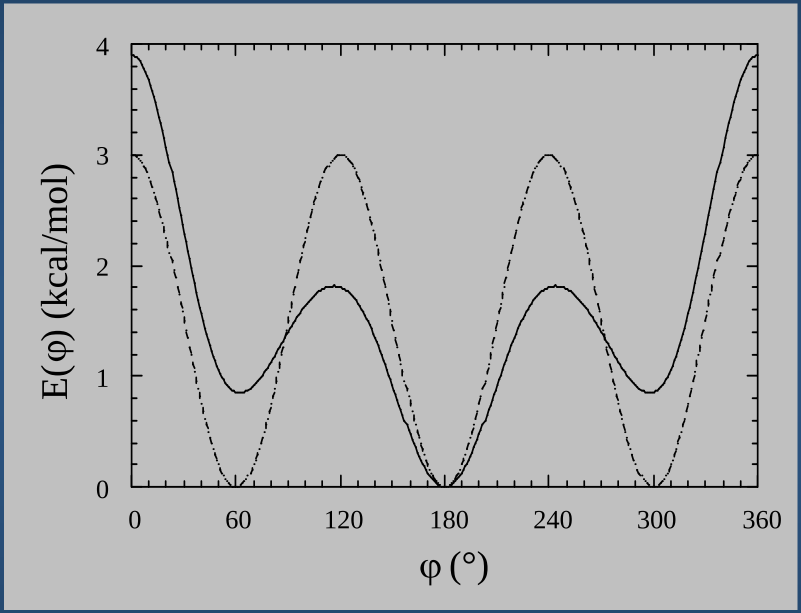
- Fluctuations around isomeric states
- Jumps between isomeric states

Most likely near native state

Definition of dihedral angles

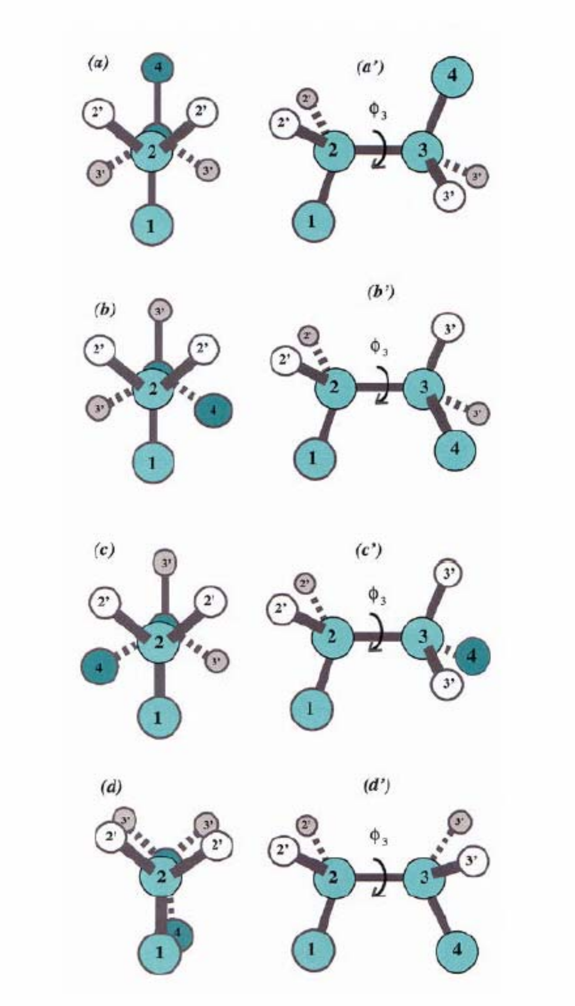


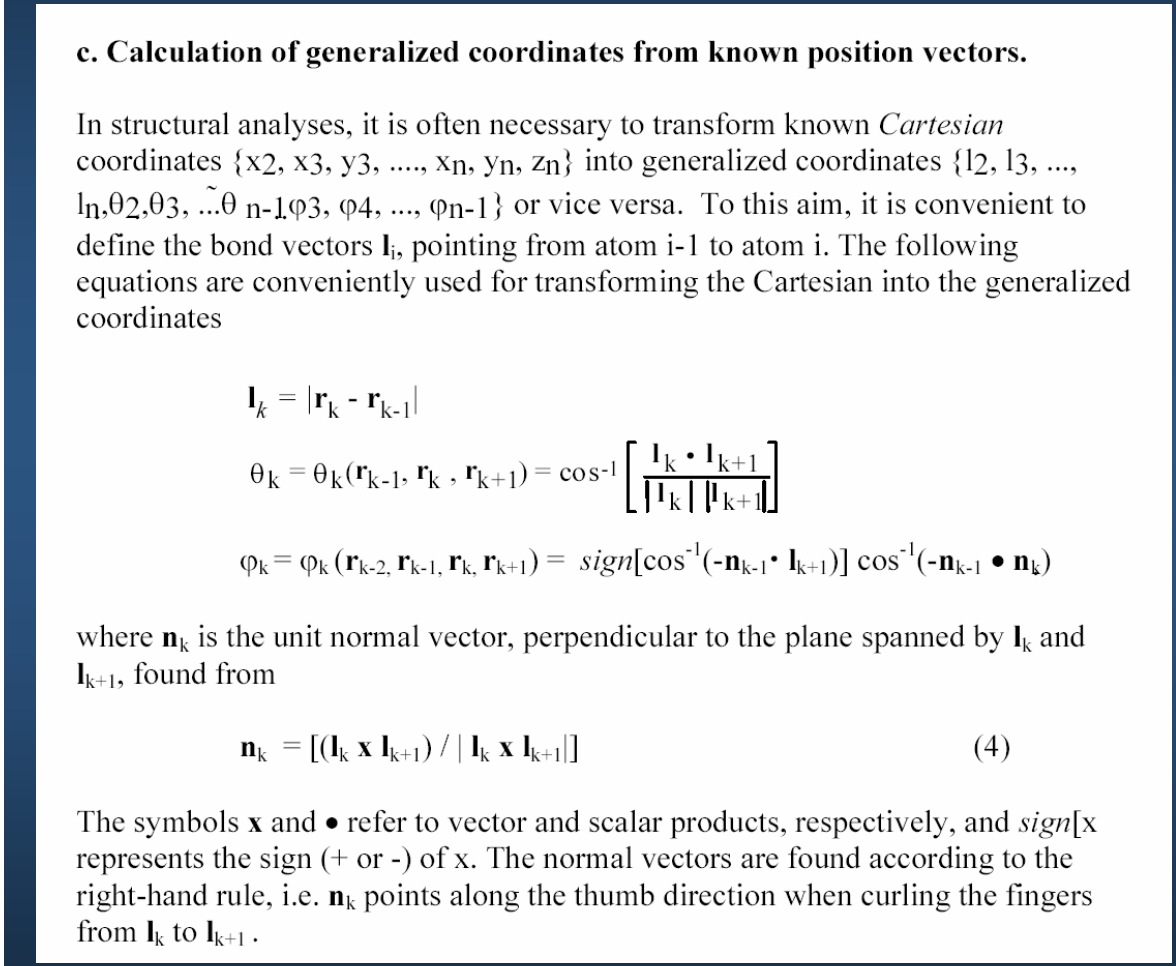
Spatial representation of the torsional mobility around the bond i+1. The torsional angle φi+1 of bond i+1 determines the position of the atom Ci+2 relative to Ci-1. C'i+2 and C"i+2 represent the positions of atom i+2, when φi+1 assumes the respective values 180° and 0°.



Rotational energy as a function of dihedral angle for a threefold symmetric torsional potential (dashed curve) and a three-state potential with a preference for the *trans* isomer (j = 180°) over the *gauche* isomers (60° and 300°) (solid curve), and the *cis* (0°) state being most unfavorable.

Rotational Isomeric States (Flory – Nobel 1974)





c. Calculation of generalized coordinates from known position vectors.

In structural analyses, it is often necessary to transform known *Cartesian* coordinates {x2, x3, y3, ..., xn, yn, zn} into generalized coordinates {l2, l3, ..., ln,θ2,θ3, ...θ n-1,φ3, φ4, ..., φn-1} or vice versa. To this aim, it is convenient to define the bond vectors $\mathbf{l}_i$, pointing from atom i-1 to atom i. The following equations are conveniently used for transforming the Cartesian into the generalized coordinates

$$\mathbf{l}_k = |\mathbf{r}_k - \mathbf{r}_{k-1}|$$

$$\theta_k = \theta_k(\mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}) = \cos^{-1}\left[\frac{\mathbf{l}_k \bullet \mathbf{l}_{k+1}}{|\mathbf{l}_k|\,|\mathbf{l}_{k+1}|}\right]$$

$$\varphi_k = \varphi_k(\mathbf{r}_{k-2}, \mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}) = \mathit{sign}[\cos^{-1}(-\mathbf{n}_{k-1} \cdot \mathbf{l}_{k+1})] \cos^{-1}(-\mathbf{n}_{k-1} \bullet \mathbf{n}_k)$$

where $\mathbf{n}_k$ is the unit normal vector, perpendicular to the plane spanned by $\mathbf{l}_k$ and $\mathbf{l}_{k+1}$, found from

$$\mathbf{n}_k = [(\mathbf{l}_k \mathbf{x} \mathbf{l}_{k+1}) / |\mathbf{l}_k \mathbf{x} \mathbf{l}_{k+1}|] \qquad (4)$$

The symbols **x** and • refer to vector and scalar products, respectively, and *sign*[x represents the sign (+ or -) of x. The normal vectors are found according to the right-hand rule, i.e. $\mathbf{n}_k$ points along the thumb direction when curling the fingers from $\mathbf{l}_k$ to $\mathbf{l}_{k+1}$.

trans → 0º ; *cis* → 180º ; *gauche* ⌖ = ⌖ 120º (Flory convention)
trans →180º ; *cis* → 0º ; *gauche* ⌖ = 60 and 300º (Bio-convention)

Bond-based coordinate systems

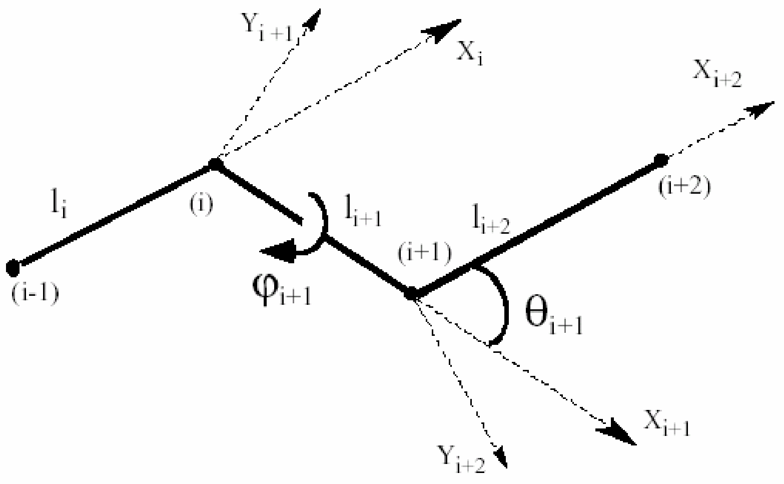


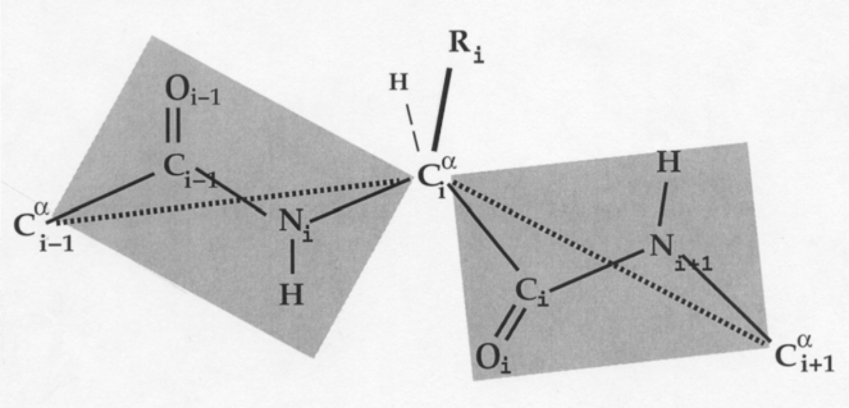
Figure 5. *Schematic representation of a chain segment of four bonds. Atomic serial indices are indicated in parentheses. The ith bond connects atoms i-1 and i along the main chain, and its torsion angle is denoted as φ_i. θ_i is the supplemental bond angle defined by bonds i and i+1. The X_{i+1} and Y_{i+1} axes of the bond-based coordinate system $X_{i+1}Y_{i+1}Z_{i+1}$ appended to the bond i+1 are shown. Y_{i+1} lies in the plane defined by bonds i and i+1, and makes an acute*

Transformation matrix between frames i+1 and i

$T_i(\theta_i, \varphi_i) =$

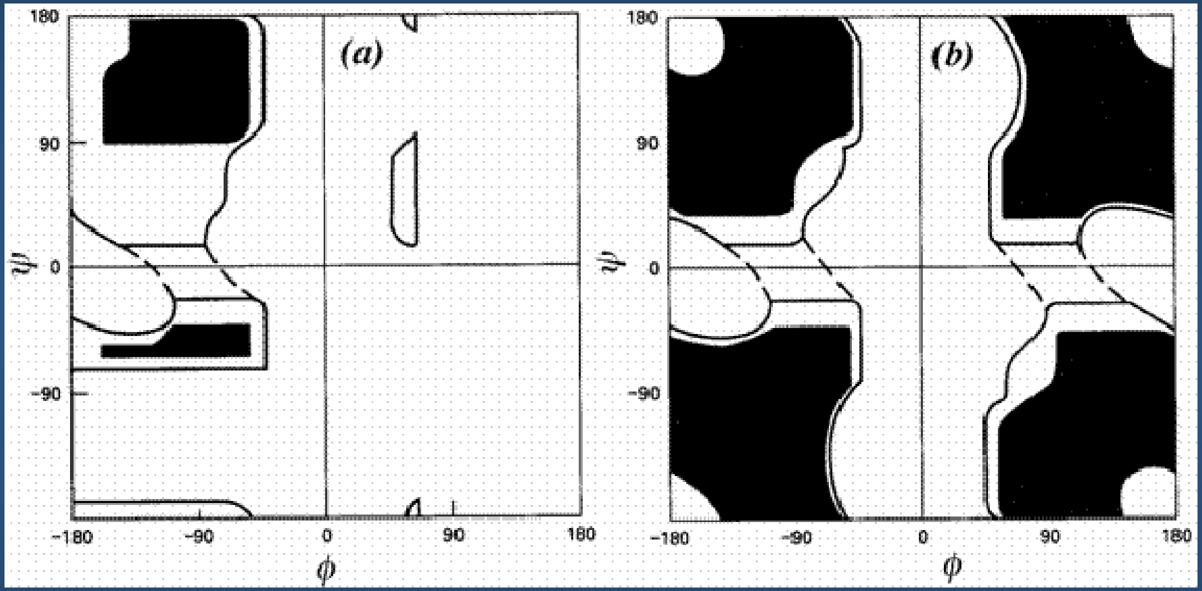
cosθi	**sinθi**	**0**
sinθi cosφi	**-cosθi cosφi**	**sinφi**
sinθi sinφi	**-cosθi sinφi**	**cosφi**

Virtual bond representation of protein backbone



Flory, PJ. Statistical Mechanics of Chain Molecules, 1969, Wiley-Interscience – Appendix B

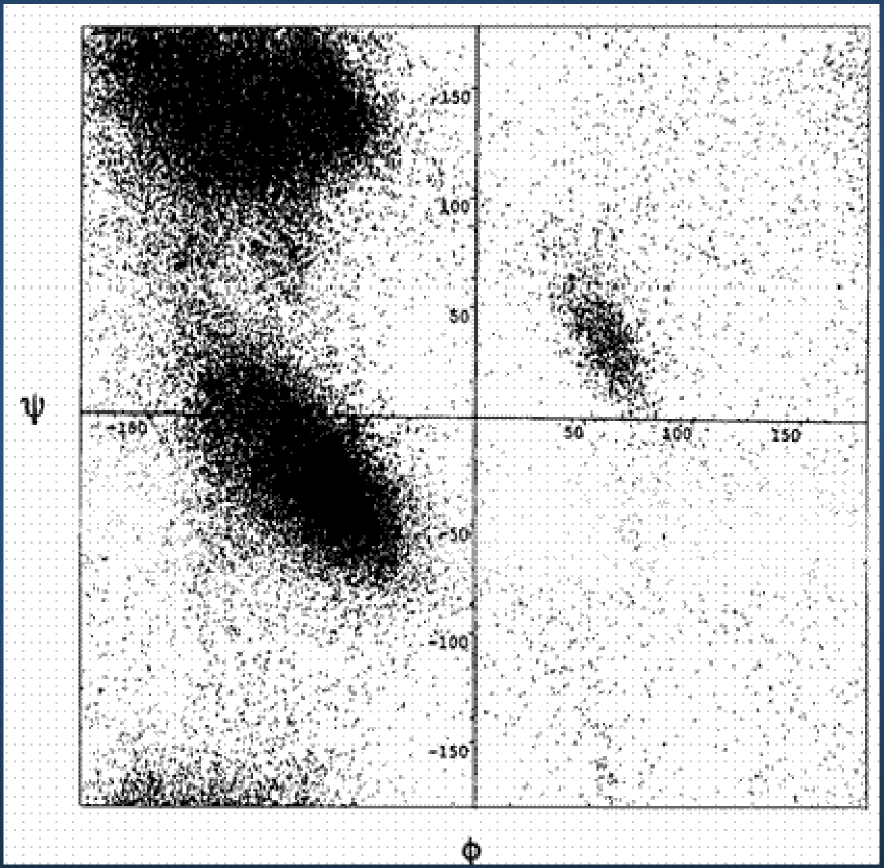
Ramachandran plots



All residues Glycine

The presence of chiral Cα atoms in Ala (and in all other amino acids) is responsible for the asymmetric distribution of dihedral angles in part (a), and the presence of Cβ excludes the portions that are accessible in Gly.

Dihedral angle distributions of database structures

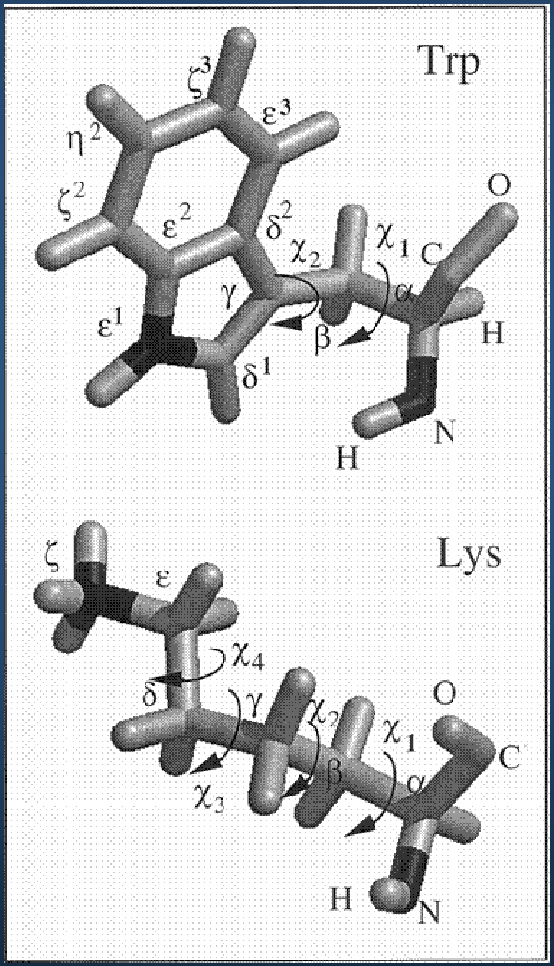


Dots represent the observed (ϕ, ψ) pairs in 310 protein structures in the Brookhaven Protein Databank (adapted from (Thornton, 1992))

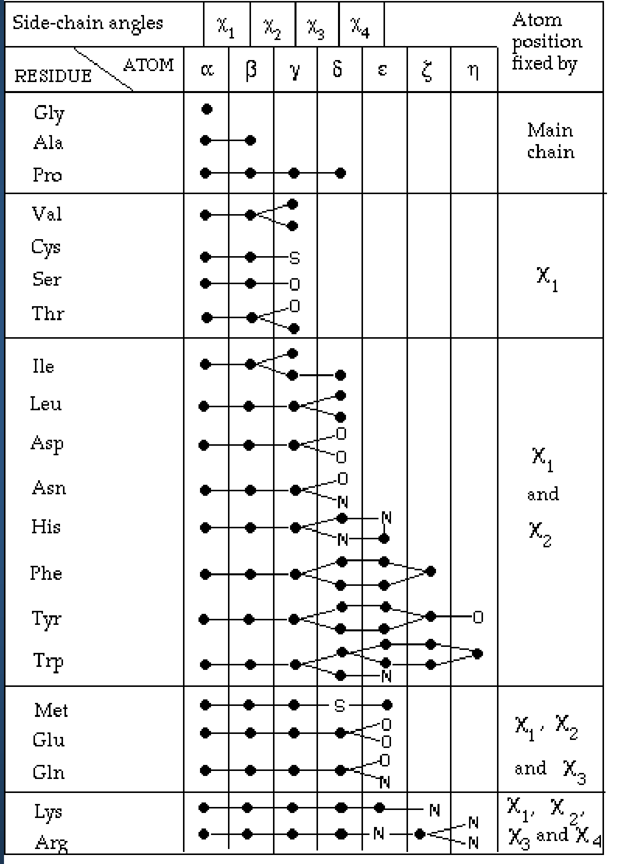
Homework 1: Passage between Cartesian coordinates and generalized coordinates

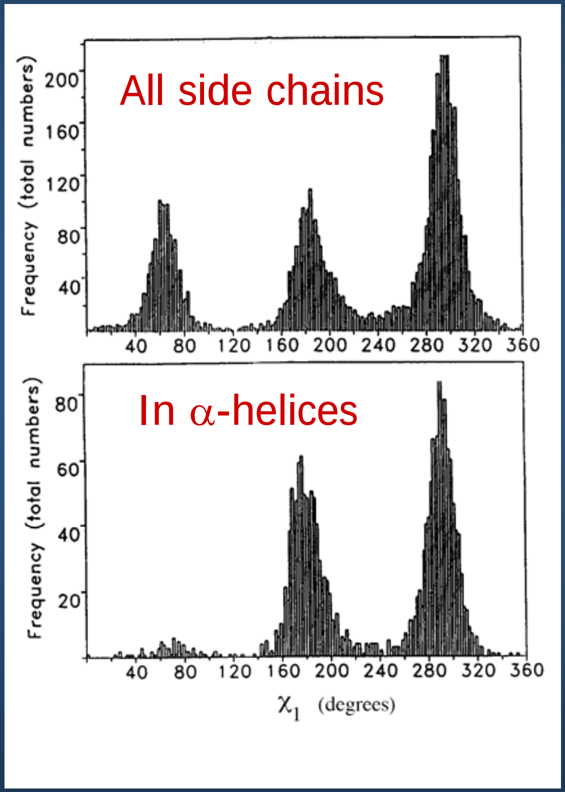
- Take a PDB file. Read the position vectors (X-, Y- and Z-coordinates – Cartesion coordinates) of the first five alpha-carbons
- Evaluate the corresponding generalized coordinates, i.e. the bond lengths l_i (i=2-5), bond angles θ_i (i=2-4), and dihedral angles ϕ_3 and ϕ_4 using the Flory convention for defining these variables.
- Using the PDB position vectors for alpha-carbons 1, 2 and 3, generate the alpha carbons 4 and 5, using the above generalized coordinates and bond-based transformation matrices. Verify that the original coordinates are reproduced.

Side chains enjoy additional degrees of freedom

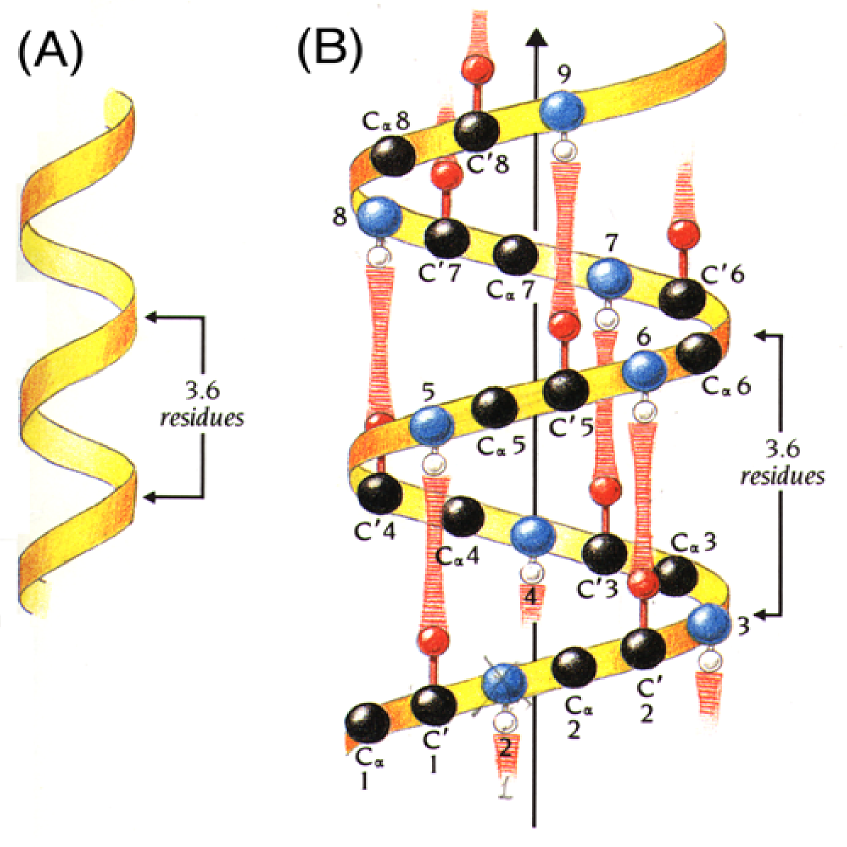


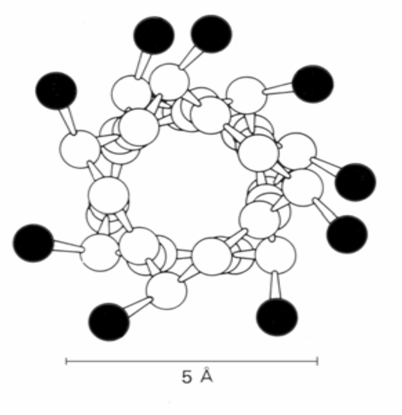
Amino acid side chains – Chi angles

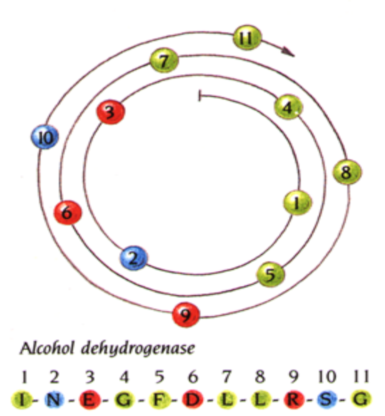




Secondary Structures: Helices and Sheets are Common Motifs

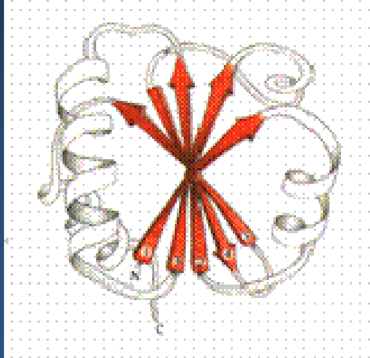


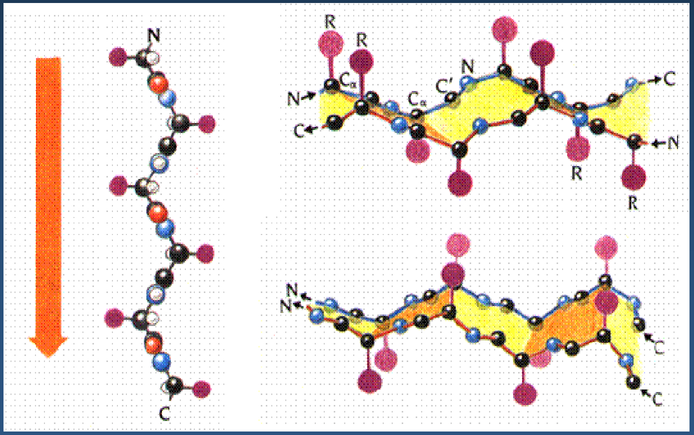


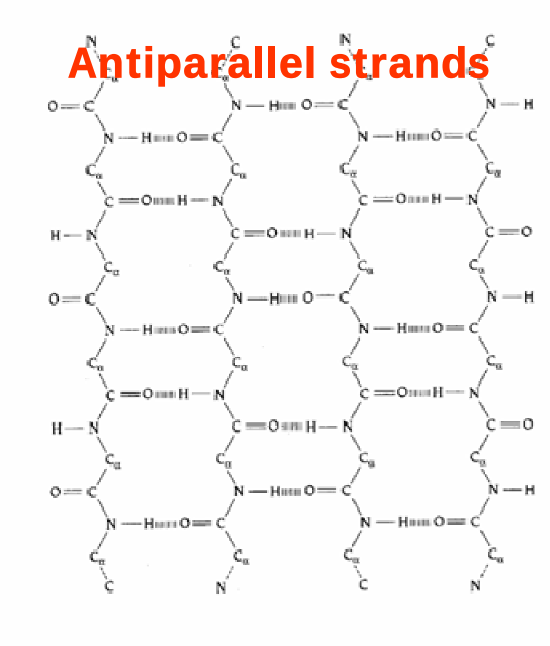


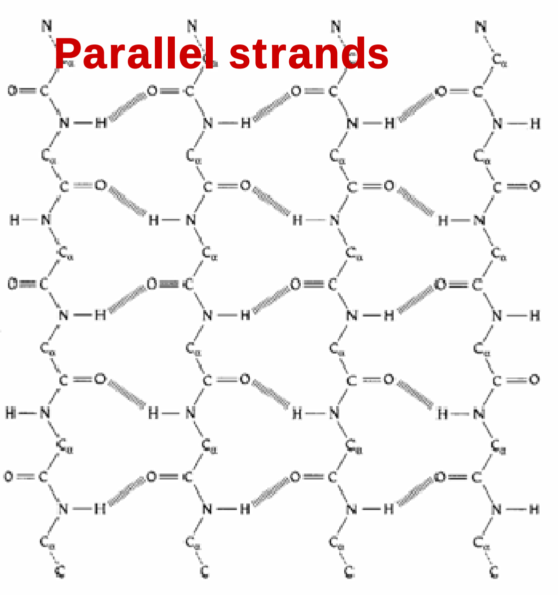
Helical wheel diagram

β-sheets: regular structures stabilized by long-range interactions

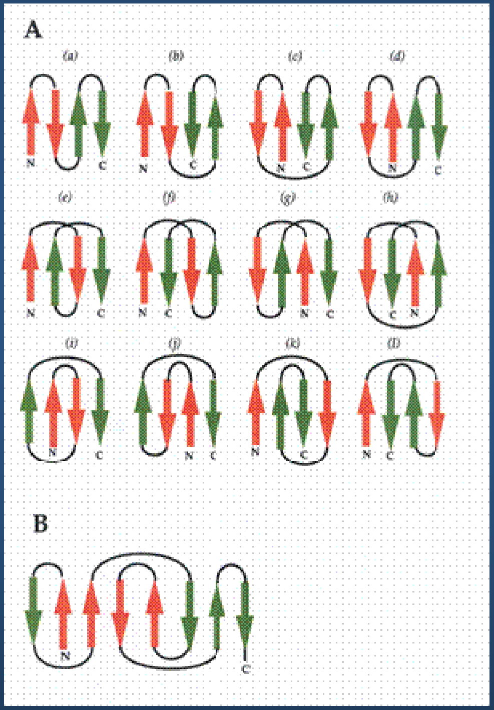


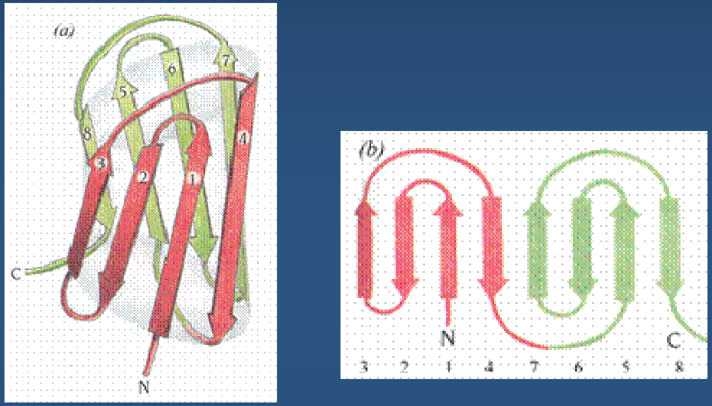



Antiparallel strands


Parallel strands

Topology diagrams for strand connections in β-sheets

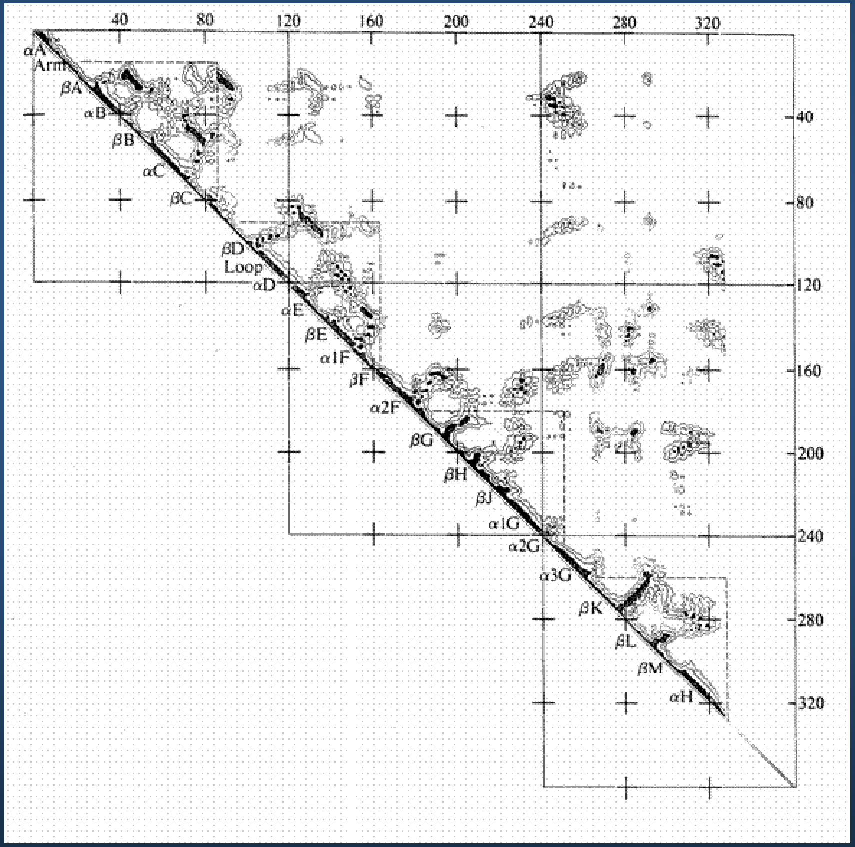


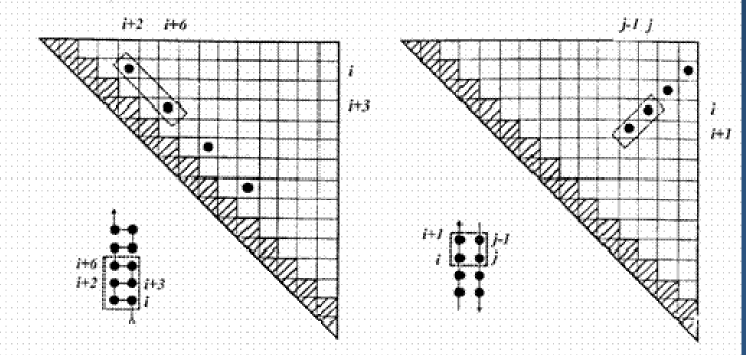


Schematic view of a β-barrel fold formed by the combination of two Greek key motifs, shown in red and green, and the topology diagram of the Greek key motifs forming the fold (adapted from Branden and Tooze, 1999)

Only those topologies where sequentially adjacent β-strands are antiparallel to each other are displayed. **(A)** 12 different ways to form a four-stranded β–sheet from two β-hairpins (red and green), if the consecutive strands 2 and 3 are assumed to be antiparallel. Not all topologies are equally probable. (j) and (l) are the most common topologies, also known as Greek key motifs; (a) is also relatively frequent; whereas (b), (c), (e), (f), (h), (i) and (k) have not been observed in known structures (Branden and Tooze, 1999).

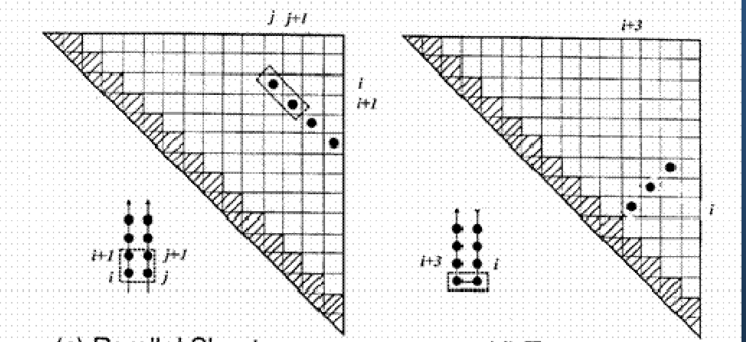
Contact Maps Describe Protein Topologies





(a) Helices

(b) Antiparallel Sheets



(c) Parallel Sheets

(d) Turns

Harmonic Oscillator Model

- Rapid movements of atoms about a valence bond
- Oscillations in bond angles
- Fluctuations around a rotational isomeric state
- Domain motions – fluctuations between open and closed forms of enzymes

Harmonic Oscillator Model

$$F = -k\,x$$

A linear motion: Force scales linearly with displacement

The corresponding **equation of motion** is of the form

$$m\,d^2x/dt^2 + k\,x = 0$$

The solution is the sinusoidal function $x = x_0 \sin(\omega t+\phi)$ where ω is the *frequency* equal to $(k/m)^{1/2}$, x_0 and ϕ are the original position and phase.

Energy of a harmonic oscillator

- Kinetic energy: $E_K = \frac{1}{2} mv^2$

where $v = dx/dt = d\,[x_0 \sin(\omega t + \phi)]/dt = x_0 \omega \cos(\omega t + \phi)$

$\rightarrow E_K = \frac{1}{2} m x_0^2 \omega^2 \cos^2(\omega t+\phi) = \frac{1}{2} m\omega^2 (x_0^2 - x^2)$

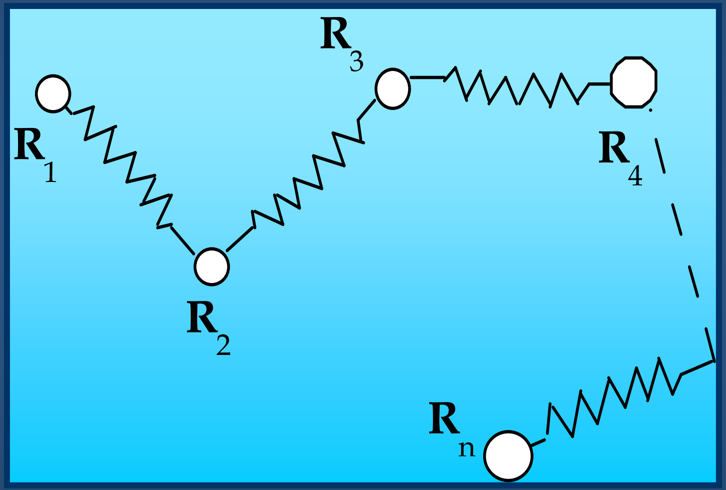
(because $x = x_0 \sin(\omega t + \phi)$ or $x^2 = x_0^2\,[1- \cos^2(\omega t+\phi)] \rightarrow x_0^2 \cos^2(\omega t+\phi) = x_0^2 - x^2$)

- Potential energy: $E_P = \frac{1}{2} kx^2$

- Total energy: $E_P + E_K = \frac{1}{2} kx_0^2$

Always fixed

Rouse chain model for macromolecules



Connectivity matrix

$$\Gamma = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & & .. & ... & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}$$

$$V_{tot} = (\gamma/2)\,[\,(\Delta R_{12})^2 + (\Delta R_{23})^2 + \ldots\ldots\ldots\, (\Delta R_{N-1,N})^2\,]$$

$$= (\gamma/2)\,[\,(\Delta R_1 - \Delta R_2)^2 + (\Delta R_2 - \Delta R_3)^2 + \ldots\ldots\ldots \tag{1}$$

Homework 2: Potential energy for a system of harmonic oscillators

(a) Using the components ΔXi, ΔYi and ΔZi of Δ**R**i, show that Eq 1 (Rouse potential) can be decomposed into three contributions, corresponding to the fluctuations along x-, y- and z-directions:

$$V_{tot} = V_X + V_Y + V_{Z.}$$

where

$$\boxed{V_X = (\gamma/2)\,[\,(\Delta X_1 - \Delta X_2)^2 + (\Delta X_2 - \Delta X_3)^2 \;+ \;........ \;(2)}$$

and similar expressions hold for Vy and Vz.

(b) Show that eq 2 can alternatively be written as

$$\boxed{V = \gamma\, ½\, \Delta X^T\, \Gamma\, \Delta X} \qquad (3)$$

where $\Delta X^T = [\Delta X_1 \;\; \Delta X_2\; \Delta X_3.....\Delta X_N]$, and ΔX is the corresponding column vector.

Hint: start from eq 3, obtain eq 2.

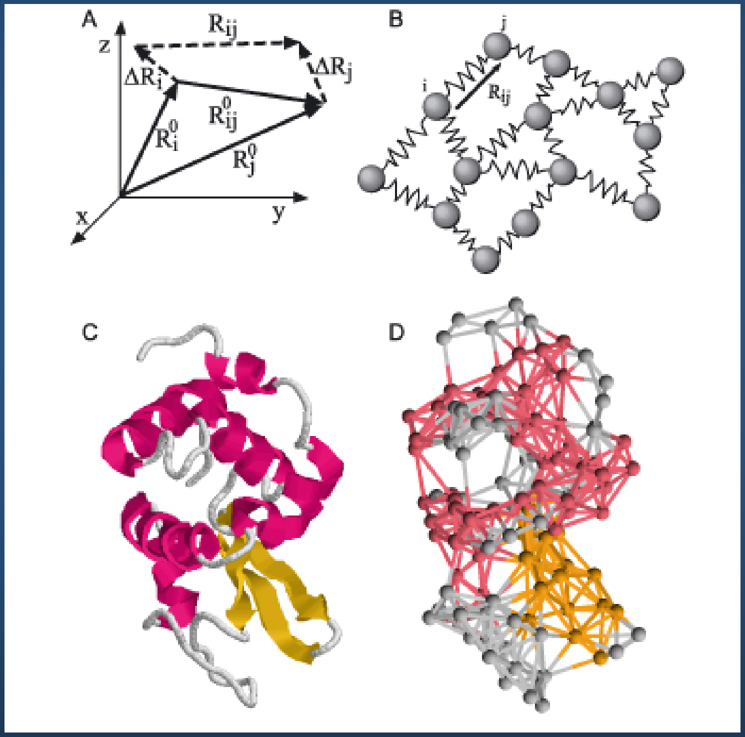
Harmonic oscillators → Gaussian distribution of fluctuations

- Consider a network formed of beads/nodes (residues or groups of residues) and springs (native contacts)

- Residues/nodes undergo Gaussian fluctuations about their mean positions – similar to the elastic network (EN) model of polymer gels (Flory)

$$W(\Delta \mathbf{R}_i) = \exp\{ -3\,(\Delta \mathbf{R}_i)^2/2\,\langle(\Delta \mathbf{R}_i)^2\rangle\}$$

Proteins can be modeled as an ensemble of harmonic oscillators



Gaussian Network Model - GNM

Molecular Movements

Physical properties of gases – a short review (Benedek & Villars, Chapter 2)

Ideal gas law: $PV_M = RT$

$$PV = NkT$$
$$PV = nRT$$

where V_M is the molar volume, T is the absolute temperature, R is the gas constant (1.987×10^{-3} kcal/mol or 8.314 J/K), k is the Boltzmann constant, N is the number of molecules, n is the number of moles = N/N_0, N_0 is the Avogadro's number.

Mean kinetic energy of a molecule of mass m and its mean-square velocity:

$$\langle \tfrac{1}{2} mv^2 \rangle = (3/2)\, kT \rightarrow \langle v^2 \rangle = (3kT/m)$$

$$v_{rms} = \langle v^2 \rangle^{1/2} = (3kT/m)^{1/2}$$

Physical kinetics – Kinetic theory of gases

Root-mean-square velocities

$$v_{rms} = \langle v^2 \rangle^{1/2} = (3kT/m)^{1/2}$$

Molecule	M (g/mol)	v_{rms} (m/s)
H_2	2	1880
O_2	32	474
Macromolecules	$10^4 - 10^6$	2.6 - 26
Viruses (e.g. tobacco mosaic virus)	$10^8 - 10^{10}$ (5×10^7 g/mol)	0.026 – 0.26 (35 cm/s)

Brownian motion
(Brown, 1827)

These numbers provide estimates on the time/length scales of **fluctuations** or Brownian motions

Equipartition law

An energy of ½ kT associated with each degree of freedom

For a diatomic molecule, there are three translational (absolute), two rotational degrees of freedom, and the mean translational energies are

$$< \frac{1}{2} m v_x^2 > = < \frac{1}{2} m v_Y^2 > = < \frac{1}{2} m v_Z^2 > = \frac{1}{2} kT$$

And the mean rotational energy is kT. For non interacting single atom molecules (ideal gases), there are only translational degrees of freedom such that the total internal energy is

$U = (3/2)kT$ and specific heat is $C_v = \partial U/\partial T = (3/2)\ k$

Random Walk

$P_N(R, L) = (1/2^N)\ N!\ /\ R!\ L!$

Probability of R steps to the right and L steps to the left in a random walk of N steps

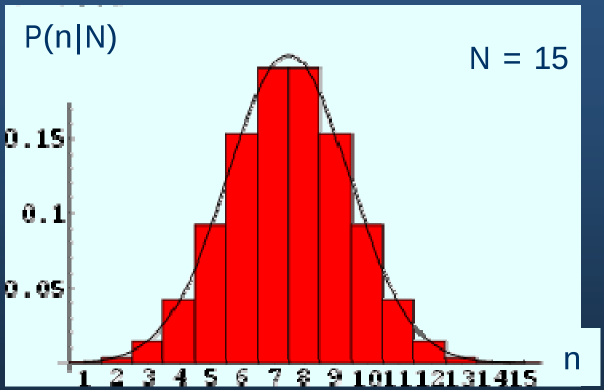
$$\boxed{\begin{matrix} R + L = N \\ R - L = m \end{matrix}} \rightarrow P_N(m) = (1/2^N)\ N!\ /([(N + m)/2]!\ [(N - m)/2]!)$$

Probability of ending up at m steps away from the origin, at the end of N steps

Binomial (or Bernoulli) Distribution

$$P_p(n|N) = \binom{N}{n} p^n q^{N-n} = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n},$$

http://mathworld.wolfram.com/BinomialDistribution.html



Properties of Binomial Distribution

Mean	**Np**
Variance	**Npq**
Standard deviation	**$(Npq)^{1/2}$**

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\bar{n})^2}{2\sigma^2}\right], = \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(n-Np)^2}{2Npq}\right].$$

Gaussian form of Bernoulli distribution

$$P_N(m) = (1/2^N)\ N!\ /\ \{[(N + m)/2]!\ [(N - m)/2]!\}$$

As m increases, the above distribution may be approximated by a continuous function

$$P_N(m) = (2/\pi N)^{1/2}\ \exp\ \{-m^2/2N\}$$

Gaussian approximation

Length of
Each step

Examples of Gaussianly distributed variables:

- Displacement (by random walk) along x-direction → $W(x) \approx \exp\ \{-x^2/2Nl^2\}$ where $m=x/l$
- Fluctuations near an equilibrium position → $W(r) \approx \exp\ \{-3(\Delta r)^2/2\langle(\Delta r)^2\rangle_0\}$
- Maxwell-Boltzmann distribution of velocities → $P(v_x) = (m/2\pi kt)^{1/2}\ \exp\ (-\tfrac{1}{2}mv_x^2/kT\}$
- Time-dependent diffusion of a particle → $P(x,t) = \sqrt{}[4\pi Dt]\ \exp(-x^2/4Dt\}$

Examples of Gaussianly distributed variables:

- Displacement (by random walk) along x-direction → $W(x) \approx \exp \{-x^2/2Nl^2\}$ where $m=x/l$
- Fluctuations near an equilibrium position → $W(r) \approx \exp \{-3(\Delta r)^2/2<(\Delta r)^2>_0\}$
- Maxwell-Boltzmann distribution of velocities → $P(v_x) = (m/2\pi kt)^{½} \exp (-½ m v_x^2/kT\}$
- Time-dependent diffusion of a particle → $P(x,t) = \sqrt{}[4\pi Dt] \exp(-x^2/4Dt\}$