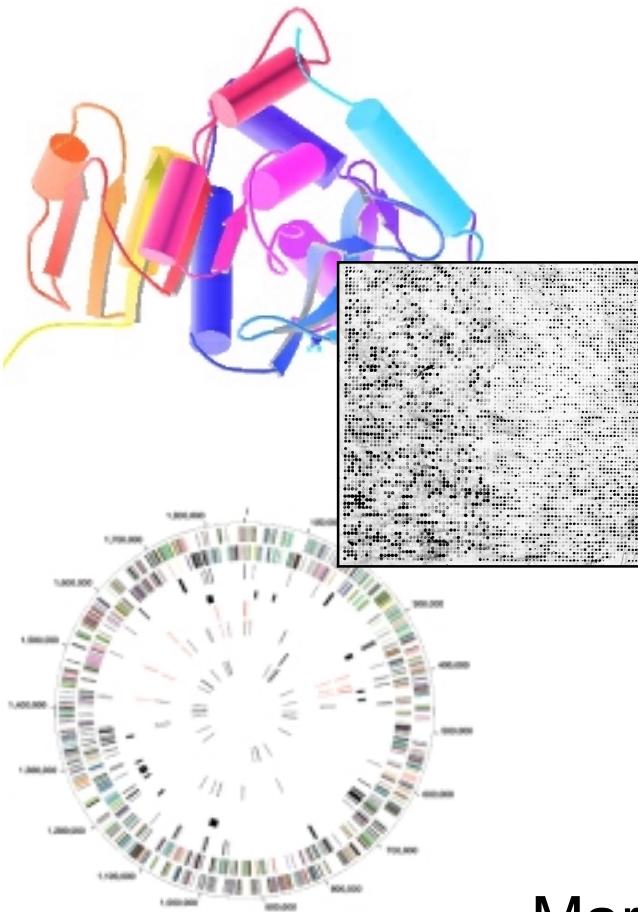


# BIOINFORMATICS

## Simulation



Mark Gerstein, Yale University  
[bioinfo.mbb.yale.edu/mgb452a](http://bioinfo.mbb.yale.edu/mgb452a)

# Overview: Electrostatics + Basic Forces

- Electrostatics
  - ◊ Polarization
  - ◊ Multipoles, dipoles
  - ◊ VDW Forces
  - ◊ Electrostatic Interactions
- Basic Forces
  - ◊ Electrical non-bonded interactions
  - ◊ bonded, fundamentally QM but treat as springs
  - ◊ Sum up the energy
- Simple Systems First

# Overview:

## Methods for the Generation and

## Analysis of Macromolecular Simulations

### 1 Simulation Methods

- ◊ Potential Functions
- ◊ Minimization
- ◊ Molecular Dynamics
- ◊ Monte Carlo
- ◊ Simulated Annealing

### 2 Types of Analysis

- ◊ liquids: RDFs, Diffusion constants
- ◊ proteins: RMS, Volumes, Surfaces

- Established Techniques (chemistry, biology, physics)
- Focus on simple systems first (liquids). Then explain how extended to proteins.

- $E$  = electric field = direction that a positive test charge would move
- $\text{Force}/q = E$
- Potential =  $W/q =$  work per unit charge =  $Fx/q = Ex$ 
  - ◊  $E = - \text{grad } \phi ; E = (d\phi/dx, d\phi/dy, d\phi/dz)$

## Electric potential, a quick review

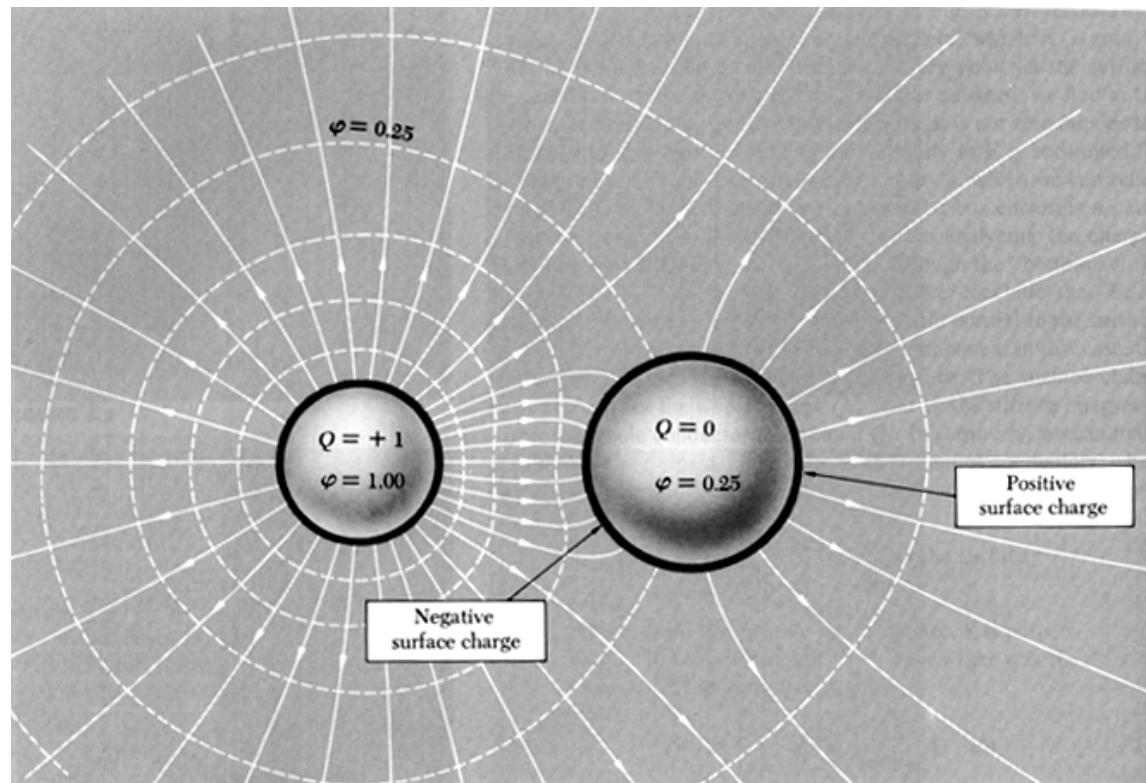


Illustration Credit: Purcell

# Maxwell's Equations

- 1st Pair (curl's)

- ◊ A changing electric field gives rise to magnetic field that circles around it & vice-versa. Electric Current also gives rise to magnetic field.  
[no discuss here]

- 2nd Pair (div's)

- ◊ Relationship of a field to sources
  - ◊ no magnetic monopoles and magnetostatics:  $\text{div } \mathbf{B} = 0$   
[no discuss here]

- All of Electrostatics in Gauss's Law!!

$$\text{curl } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$$

$$\text{curl } \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{J}$$

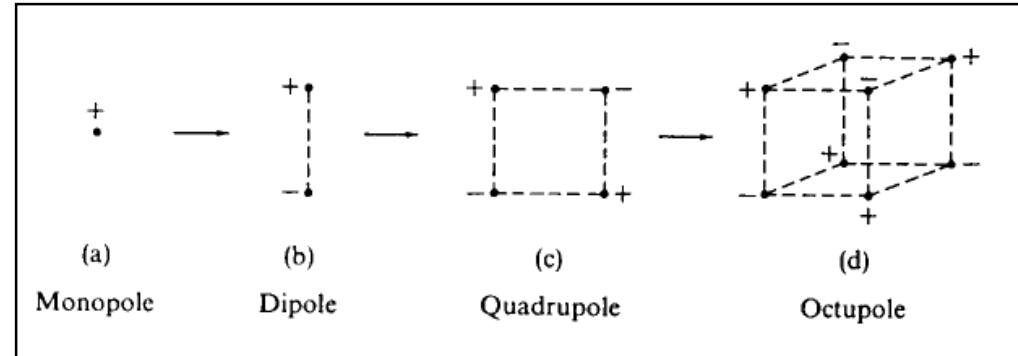
$$\text{div } \mathbf{E} = 4\pi\rho$$

$$\text{div } \mathbf{B} = 0$$

cgs (not mks) units above

# Multipole Expansion

- Routinely done when an atom's charge distribution is replaced by a point charge or a point charge and a dipole
  - ◊ Ignore above dipole here
  - ◊ Harmonic expansion of pot.
- Only applicable far from the charge distribution
  - ◊ Helix Dipole not meaningful close-by
- Terms drop off faster with distance



$$\Phi(\mathbf{x}) = \frac{q}{r} + \frac{\mathbf{p} \cdot \mathbf{x}}{r^3} + \frac{1}{2} \sum_{i,j} Q_{ij} \frac{x_i x_j}{r^5} + \dots$$

$$\Phi(\mathbf{x}) = \frac{K_1 q}{r} + \frac{K_2 q}{r^2} + \frac{K_3 q}{r^3} + \dots$$

Replace continuous charge distribution with point moments: charge (monopole) + dipole + quadrupole + octupole + ...

# Gauss' Law: Electrostatics

- $\operatorname{div} \mathbf{E} = 4\pi\rho$
- Coulomb's Law
  - ◊  $\int \operatorname{div} \mathbf{E} dV = \int 4\pi\rho dV$
  - ◊  $\int \mathbf{E} \cdot d\mathbf{A} = \int 4\pi\rho dV$  [Divergence thm.]
  - ◊ Assume spherically symmetrical charge distribution
  - ◊  $E(4\pi r^2) = 4\pi Q \implies E = Q/r^2$
  - ◊  $U = -Q/r$  [assuming a zero at inf.]
- Equations for the Potential Based on the Charge in a Region plus Boundary Conditions
  - ◊  $\operatorname{div} \operatorname{grad} U = 4\pi\rho$
  - ◊  $\nabla^2 U = 4\pi\rho$  [poisson's equation]
  - ◊  $\nabla^2 U = 0$  [Laplace's equation]

- $\phi(r, \theta) = -q/R_1 + q/R_2$   
 $\diamond \phi(r, \theta) = q(R_1 - R_2)/R_1 R_2$
- If  $r$  is very much larger than  $L$ 
  - $\diamond$  Vectors essentially parallel, like single-slit
  - $\diamond R_1 R_2 = r^2$
  - $\diamond R_2 - R_1 = 2L \cos \theta$
  - $\diamond q(R_2 - R_1) = 2Lq \cos \theta = p \cos \theta = \mathbf{p} \cdot \mathbf{r}/|r|$
  - $\diamond \mathbf{p}$  = dipole moment vector  
 $= [\text{charge}][\text{separation}]$   
in direction from neg. to positive charge
- $\phi(r, \theta) = p \cos \theta / r^2$ 
  - $\diamond E = \text{grad } \phi(r, \theta) \sim 1/r^3$  with a complex angular dependence
- Monopole is  $1/r$ , which dominates over dipole ( $1/r^2$ ), dipole dominates quadrupole

## Dipole Derivation

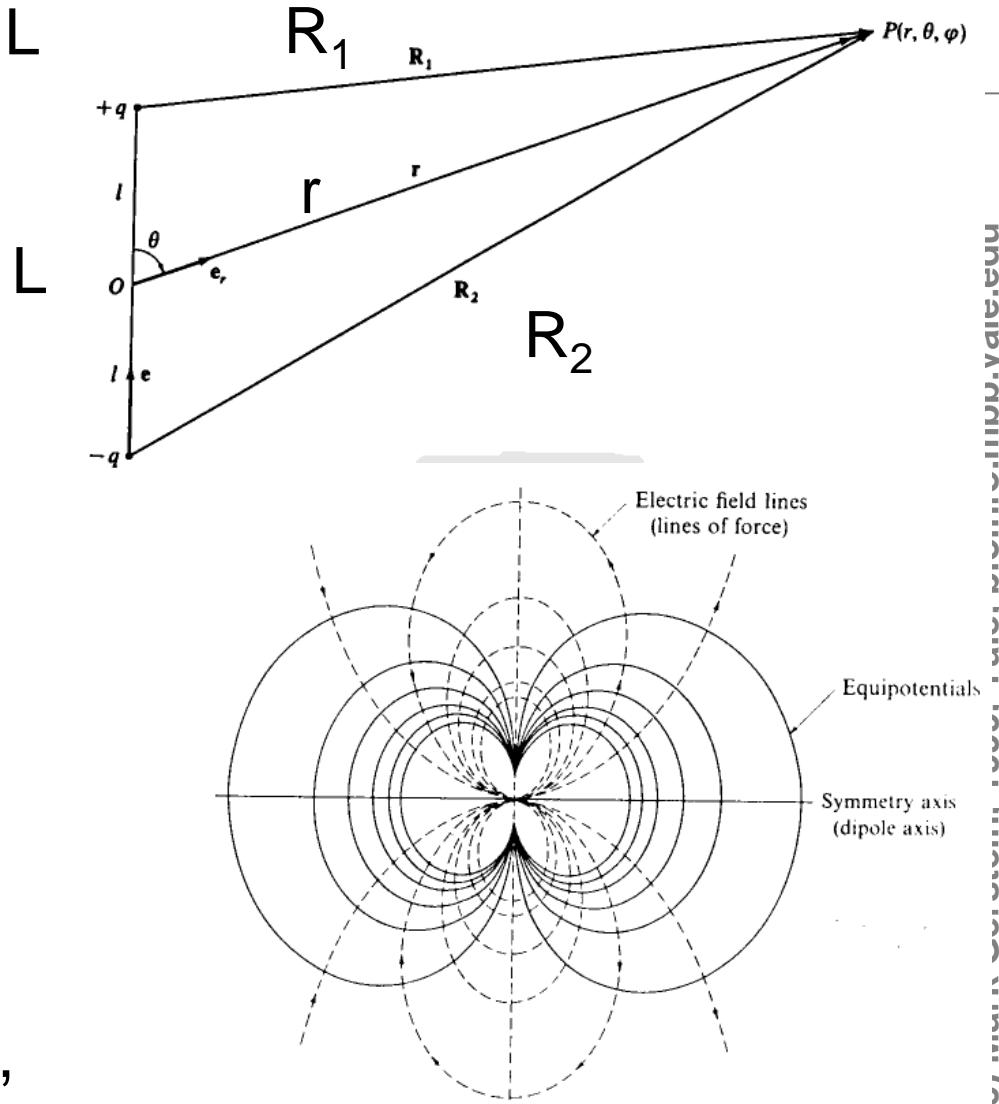
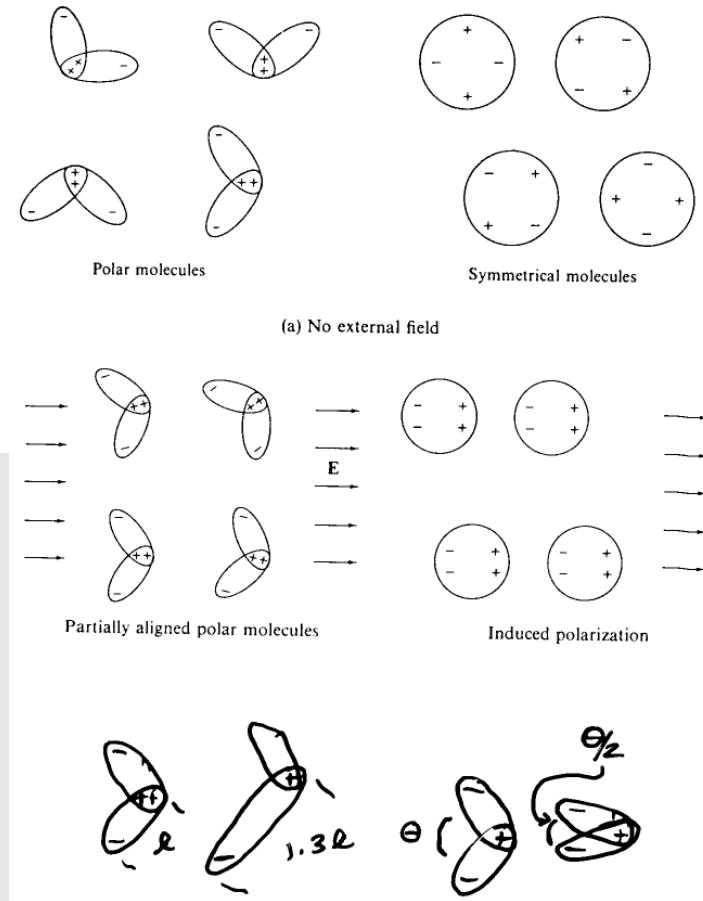
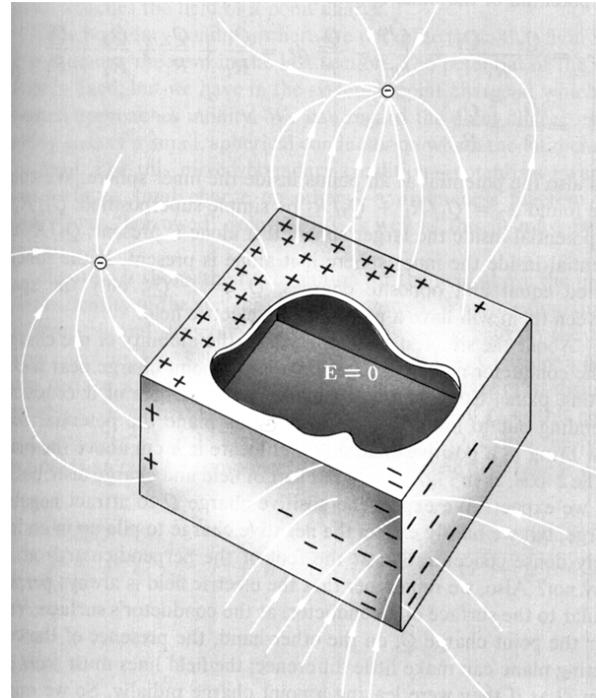


Illustration Credit: Marion & Heald

# Polarization



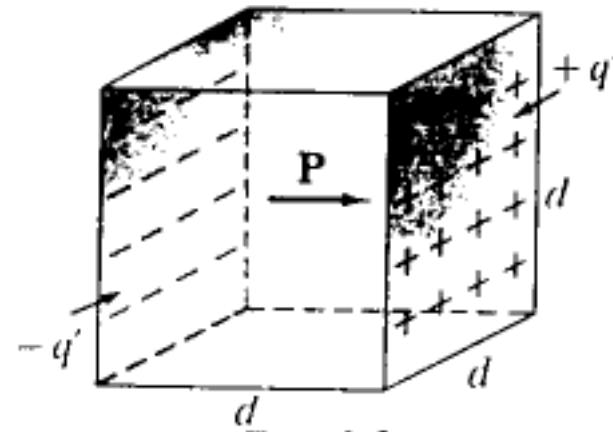
- Charge shifts to resist field
  - ◊ Accomplished perfectly in conductor -- surface charge, no field inside
  - ◊ Insulators partially accommodate via induced dipoles
- Induced dipole
  - ◊ charge/ion movement (slowest)
  - ◊ dipole reorient
  - ◊ molecular distort (bond length and angle)
  - ◊ electronic (fastest)

Illustration Credit: Purcell, Marion & Heald

$$q'd = Pd^3$$

# Dielectric const.

- Macro manifestation of polarization
- Values  
(measured in debye)
  - ◊ Air, 1
  - ◊ Water, 80
  - ◊ Paraffin Wax, 2
  - ◊ Methanol, 33
  - ◊ Non-polar protein, 2
  - ◊ Polar protein, 4
- High-frequency
  - ◊ water re-orient, 1ps
  - ◊ bond, angle stretch
  - ◊ electronic, related to index of refraction



- $P = \alpha E$   
 $P$  = dipole moment per unit volume
- $\alpha$  = electric susceptibility
- $\alpha = (\epsilon - 1)/4\pi$
- $\epsilon$  = dielectric const.
- Effective Field Inside Reduced by Polarization

# Polarity vs. Polarizability

From Sharp (1999): “Application of a classical electrostatic view to macromolecular electrostatics involves a number of useful concepts that describe the physical behavior. It should first be recognized that the potential at a particular charged atom  $i$  includes three physically distinct contributions. **The first is the direct or Coulombic potential of  $j$  at  $i$ . The second is the potential at  $i$  from the polarization (from molecule, water and ionic) induced by  $j$ . This is often referred to as the screening potential, since it opposes the direct, Coulombic potential. The third arises from the polarization induced by  $i$  itself. This is often referred to as the reaction or self potential**, and if solvent is involved, as the solvation potential. When using models which apply the concept of a dielectric constant (a measure of polarizability) to a macromolecule, **it is important to distinguish between polarity and polarizability.** Briefly, polarity may be thought of as describing the density of charged and dipolar groups in a particular region. Polarizability, by contrast, refers to the *potential* for reorganizing charges, orienting dipoles and inducing dipoles. Thus polarizability depends both on the polarity and the freedom of dipoles to reorganize in response to an applied electric field. When a protein is folding, or undergoing a large conformational rearrangement, the peptide groups may be quite free to reorient. In the folded protein these may become spatially organized so as to stabilize another charge or dipole, creating a region with high polarity, but with low polarizability, since there is much less ability to reorient the dipolar groups in response to a new charge or dipole without significant disruption of the structure. Thus, while there is still some discussion about the value and applicability of a protein dielectric constant, it is generally agreed that the interior of a macromolecule is a low polarizable environment compared to solvent. This difference in polarizability has a significant effect on the potential distribution.”

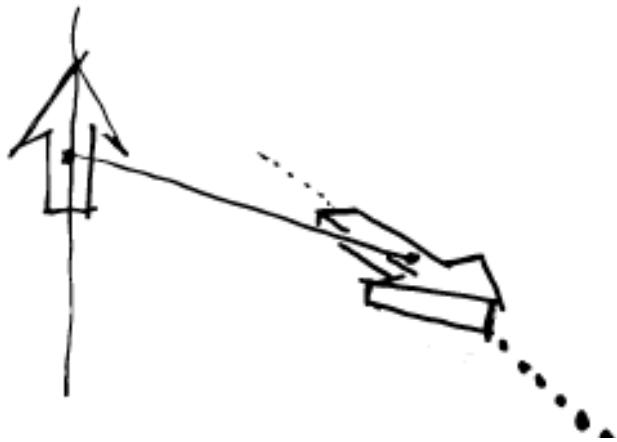
# VDW Forces:

## Start by

### Deriving

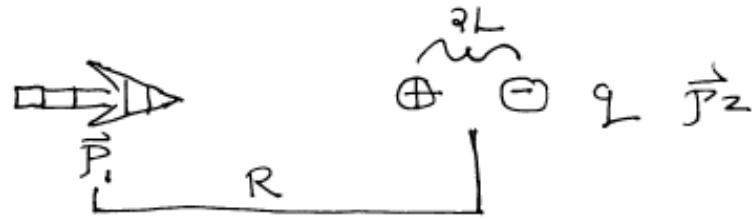
#### Dipole-Dipole

#### Energy



Interaction energy  
of a pair of dipoles  
is a complex function  
of two angles ( $\theta, \psi$ )

Simplify. Focus on Formula  
for Parallel Dipoles



$$V(R) = -q_1 p_1 \left( \frac{1}{(R-L)^2} - \frac{1}{(R+L)^2} \right)$$
$$= +q_1 p_1 \left( \frac{4RL}{(R+L)^2(R-L)^2} \right)$$

IF  $R \gg L$

$$V(R) \doteq +2 p_1 p_2 \frac{1}{R^3}$$

PARALLEL DIPOLES

IN GENERAL,  $V = \frac{C p_1 p_2}{R^3} f$

WHERE  $f$  IS A FUNCTION OF ORIENTATION  
ANGLES  $\theta$  &  $\psi$  —  $f(\theta, \psi)$

# Average Dipole- Dipole Interaction Energy

- Multiplication of dipole-dipole energy ( $1/r^3$ ) and Boltz. Factor (~dipole-dipole energy) gives ( $1/r^6$ )

AVERAGE INTERACTION ENERGY OVER ORIENTATIONS

$$\begin{aligned} \langle V(R, \theta, \psi) \rangle_{\theta, \psi} &= \langle V \rangle_{ori} \\ &= \left\langle \frac{C_{P_1 P_2}}{R^3} f(\theta, \psi) \omega(R, \theta, \psi) \right\rangle_{ori} \end{aligned}$$

$\omega$  = AMOUNT TIME SPENT AT A PARTICULAR ORIENTATION = BOLTZMANN Factor  
 $= \exp(-\frac{\langle V(R, \theta, \psi) \rangle / kT}{\text{dipole interaction energy}})$

since  $V \ll kT$ ,  
 $\omega = 1 - \frac{V}{kT} + \dots$ ,  $V = \frac{C_{P_1 P_2}}{R^3} f$

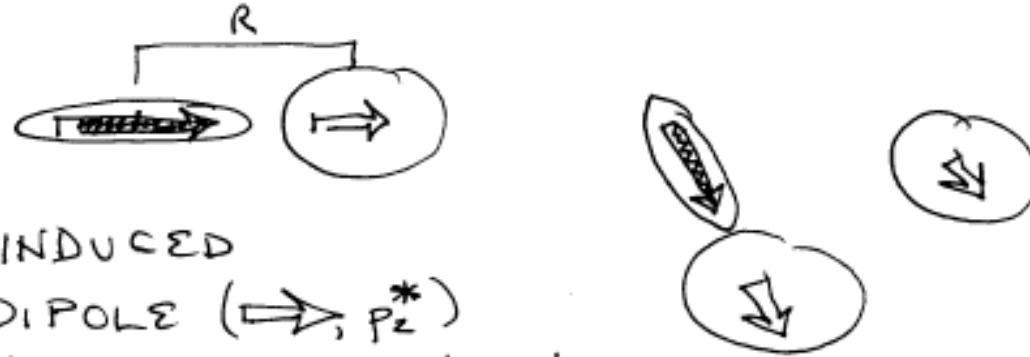
Thus,

$$\begin{aligned} \langle V \rangle_{ori} &= \left\langle \frac{C_{P_1 P_2}}{R^3} f(\theta, \psi) \left( 1 - \frac{C_{P_1 P_2}}{R^3} f(\theta, \psi) \right) \right\rangle \\ &= \frac{C_{P_1 P_2}}{R^3} \left( \langle f \rangle - \left\langle \frac{f^2 C_{P_1 P_2}}{R^3} \right\rangle \right) \\ &= -\frac{C_{P_1}^2 C_{P_2}^2}{R^6} \langle f^2 \rangle \quad \langle f^2 \rangle \propto [0, 4] \sim \frac{2}{3} \end{aligned}$$

Thus,  $\langle V \rangle_{ori} = -\frac{C'}{R^6}$

# Dipole-induced dipole Energy

- Multiplication of dipole-dipole energy ( $1/r^3$ ) and amount of induced dipole ( $1/r^3$ ) gives ( $1/r^6$ )



INDUCED  
DIPOLE ( $\Rightarrow, p_2^*$ )  
is always parallel to  
permanent dipole ( $\Rightarrow, p_1$ )

$$\vec{p}_2^* = \alpha \vec{E}$$

$$\vec{E}_{\text{dipole}} = \nabla \frac{\vec{p}_1 \cdot \hat{r}}{R^2} = -\frac{2\vec{p}_1}{R^3}$$

Using parallel dipole formula above,

$$V(R) = \frac{2\vec{p}_1 \cdot \vec{p}_2^*}{R^3} = -\frac{4\vec{p}_1^2 \alpha}{R^6} = -\frac{C}{R^6}$$

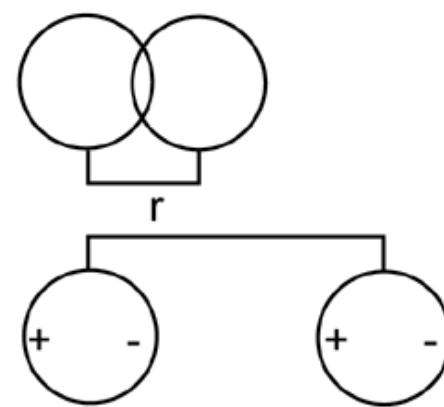
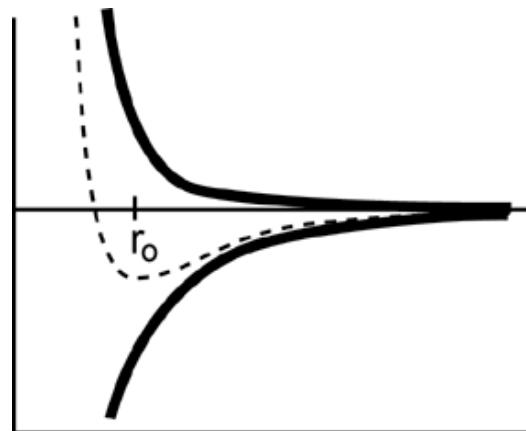
# VDW Forces:

## Induced dipole-induced dipole

- Too complex to derive induced-dipole-induced dipole formula, but it has essential ingredients of dipole-dipole and dipole-induced dipole calculation, giving an **attractive**  $1/r^6$  dependence.
  - ◊ London Forces
- Thus, total dipole cohesive force for molecular system is the sum of three  $1/r^6$  terms.
- Repulsive forces result from electron overlap.
  - ◊ Usually modeled as  $A/r^{12}$  term. Also one can use  $\exp(-Cr)$ .
- VDW forces:  $V(r) = A/r^{12} - B/r^6 = 4\epsilon((R/r)^{12} - (R/r)^6)$ 
  - ◊  $\epsilon \sim .2$  kcal/mole,  $R \sim 3.5$  Å,  $V \sim .1$  kcal/mole [favorable]

# Packing ~ VDW force

- Longer-range isotropic attractive tail provides general cohesion
- Shorter-ranged repulsion determines detailed geometry of interaction
- Billiard Ball model, WCA Theory



Electron  
Overlap  
Repulsion

$$U = \varepsilon \left( \frac{r_0}{r} \right)^{12}$$

Dispersion  
Attraction

$$U = -4\varepsilon \left( \frac{r_0}{r} \right)^6$$

# Close-packing is Default

- No tight packing when highly directional interactions (such as H-bonds) need to be satisfied
- Packing spheres (.74), hexagonal
- Water (~.35), “Open” tetrahedral, H-bonds

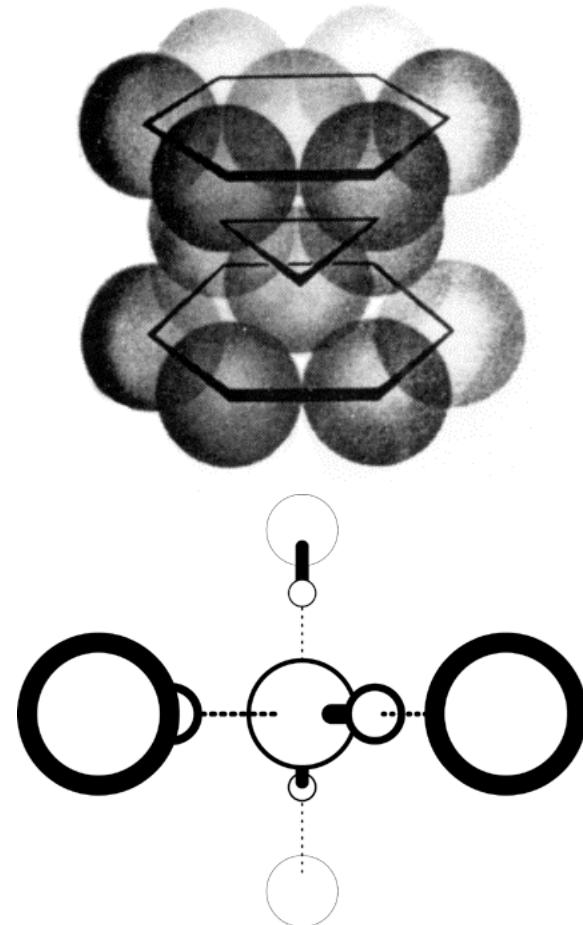
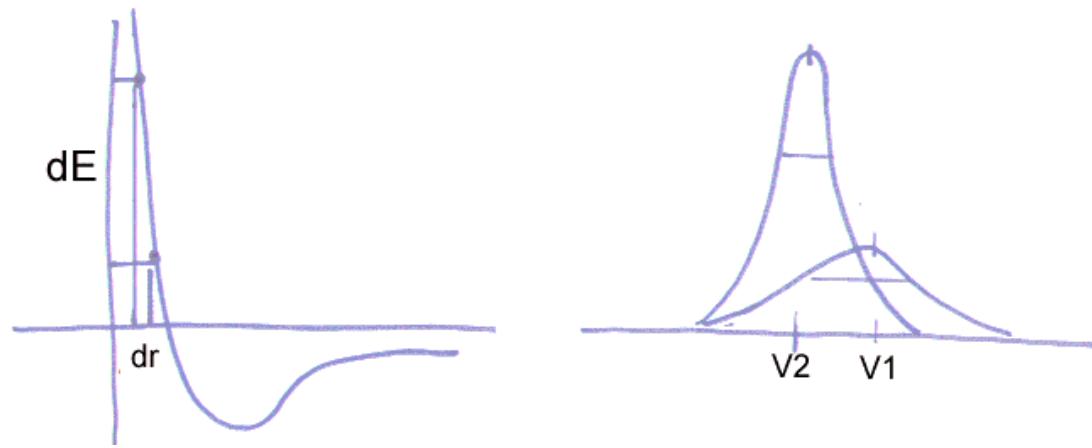


Illustration Credit: Atkins

# Small Packing Changes Significant

- Exponential dependence
- Bounded within a range of 0.5 (.8 and .3)
- Many observations in standard volumes gives small error about the mean ( $SD/\sqrt{N}$ )



| atom  | $\epsilon$<br>(kJ/mole) | $\sigma$<br>(Å) | charge<br>(electrons) |
|---|-------------------------|-----------------|-----------------------|
| carbonyl carbon                               | 0.5023                  | 3.7418          | 0.550                 |
| $\alpha$ -carbon (incorporating 1 hydrogen)   | 0.2034                  | 4.2140          | 0.100                 |
| $\beta$ -carbon (incorporating 3 hydrogens)   | 0.7581                  | 3.8576          | 0.000                 |
| amide nitrogen                                | 0.9979                  | 2.8509          | -0.350                |
| amide hydrogen                                | 0.2085                  | 1.4254          | 0.250                 |
| carbonyl oxygen                               | 0.6660                  | 2.8509          | -0.550                |
| water oxygen in interactions with the helix   | 0.6660                  | 2.8509          | -0.834                |
| water hydrogen in interactions with the helix | 0.2085                  | 1.4254          | 0.417                 |
| water O in interactions with other waters     | 0.6367                  | 3.1506          | -0.834                |
| water H in interactions with other waters     | 0.0000                  | 0.0000          | 0.417                 |

# Different Sets of Radii

Despite sensitivity of VDW radius and  $r_0$  parameter there is considerable disagreement!

| Atom Type & Symbol            | Bondi<br>1968      | Lee<br>&<br>Richards<br>1971 | Shrake<br>&<br>Rupley<br>1973 | Richards<br>1974 | Chothia<br>1975 | Rich-<br>mond &<br>Richards<br>1978 | Gelin<br>&<br>Karplus<br>1979 | Dunfield<br>et al.<br>1979 | ENCAD<br>derived<br>1995 | CHARMM<br>derived<br>1995 | Tsai<br>et al.<br>1998 |
|-------------------------------|--------------------|------------------------------|-------------------------------|------------------|-----------------|-------------------------------------|-------------------------------|----------------------------|--------------------------|---------------------------|------------------------|
| -CH <sub>3</sub>              | Aliphatic, methyl  | 2.00                         | 1.80                          | 2.00             | 2.00            | 1.87                                | 1.90                          | 1.95                       | 2.13                     | 1.82                      | 1.88                   |
| -CH <sub>2</sub> -            | Aliphatic, methyl  | 2.00                         | 1.80                          | 2.00             | 2.00            | 1.87                                | 1.90                          | 1.90                       | 2.23                     | 1.82                      | 1.88                   |
| >CH-                          | Aliphatic, CH      | -                            | 1.70                          | 2.00             | 2.00            | 1.87                                | 1.90                          | 1.85                       | 2.38                     | 1.82                      | 1.88                   |
| =CH                           | Aromatic, CH       | -                            | 1.80                          | 1.85             | *               | 1.76                                | 1.70                          | 1.90                       | 2.10                     | 1.74                      | 1.80                   |
| >C=                           | Trigonal, aromatic | 1.74                         | 1.80                          | *                | 1.70            | 1.76                                | 1.70                          | 1.80                       | 1.85                     | 1.74                      | 1.80                   |
| -NH <sub>3</sub> <sup>+</sup> | Amino, protonated  | -                            | 1.80                          | 1.50             | 2.00            | 1.50                                | 0.70                          | 1.75                       |                          | 1.68                      | 1.40                   |
| -NH <sub>2</sub>              | Amino or amide     | 1.75                         | 1.80                          | 1.50             | -               | 1.65                                | 1.70                          | 1.70                       |                          | 1.68                      | 1.40                   |
| >NH                           | Peptide, NH or N   | 1.65                         | 1.52                          | 1.40             | 1.70            | 1.65                                | 1.70                          | 1.65                       | 1.75                     | 1.68                      | 1.40                   |
| =O                            | Carbonyl Oxygen    | 1.50                         | 1.80                          | 1.40             | 1.40            | 1.40                                | 1.40                          | 1.60                       | 1.56                     | 1.34                      | 1.38                   |
| -OH                           | Alcoholic hydroxyl | -                            | 1.80                          | 1.40             | 1.60            | 1.40                                | 1.40                          | 1.70                       |                          | 1.54                      | 1.53                   |
| -OM                           | Carboxyl Oxygen    | -                            | 1.80                          | 1.89             | 1.50            | 1.40                                | 1.40                          | 1.60                       | 1.62                     | 1.34                      | 1.41                   |
| -SH                           | Sulphydryl         | -                            | 1.80                          | 1.85             | -               | 1.85                                | 1.80                          | 1.90                       |                          | 1.82                      | 1.56                   |
| -S-                           | Thioether or -S-S- | 1.80                         | -                             | -                | 1.80            | 1.85                                | 1.80                          | 1.90                       | 2.08                     | 1.82                      | 1.56                   |

# Molecular Mechanics: Simple electrostatics

- $U = kqQ/r$
- Molecular mechanics uses partial unpaired charges with monopole
  - ◊ usually no dipole
  - ◊ e.g. water has apx. -.8 on O and +.4 on Hs
  - ◊ However, normally only use monopoles for unpaired charges (on charged atoms, asp O)
- Longest-range force
  - ◊ Truncation? Smoothing

| atom  |  | $\epsilon$<br>(kJ/mole) | $\sigma$<br>(Å) | charge<br>(electrons) |
|---|--|-------------------------|-----------------|-----------------------|
| carbonyl carbon                               |  | 0.5023                  | 3.7418          | 0.550                 |
| $\alpha$ -carbon (incorporating 1 hydrogen)   |  | 0.2034                  | 4.2140          | 0.100                 |
| $\beta$ -carbon (incorporating 3 hydrogens)   |  | 0.7581                  | 3.8576          | 0.000                 |
| amide nitrogen                                |  | 0.9979                  | 2.8509          | -0.350                |
| amide hydrogen                                |  | 0.2085                  | 1.4254          | 0.250                 |
| carbonyl oxygen                               |  | 0.6660                  | 2.8509          | -0.550                |
| water oxygen in interactions with the helix   |  | 0.6660                  | 2.8509          | -0.834                |
| water hydrogen in interactions with the helix |  | 0.2085                  | 1.4254          | 0.417                 |
| water O in interactions with other waters     |  | 0.6367                  | 3.1506          | -0.834                |
| water H in interactions with other waters     |  | 0.0000                  | 0.0000          | 0.417                 |

# H-bonds subsumed by electrostatic interactions

- Naturally arise from partial charges
  - ◊ normally arise from partial charge
- Linear geometry
- Were explicit springs in older models

Illustration Credit: Taylor & Kennard (1984)

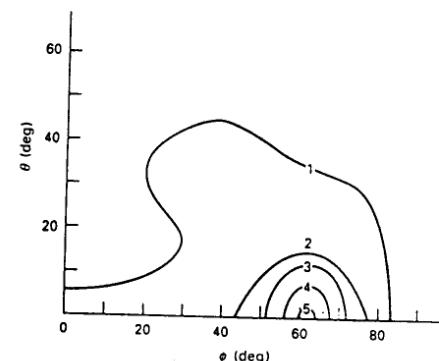
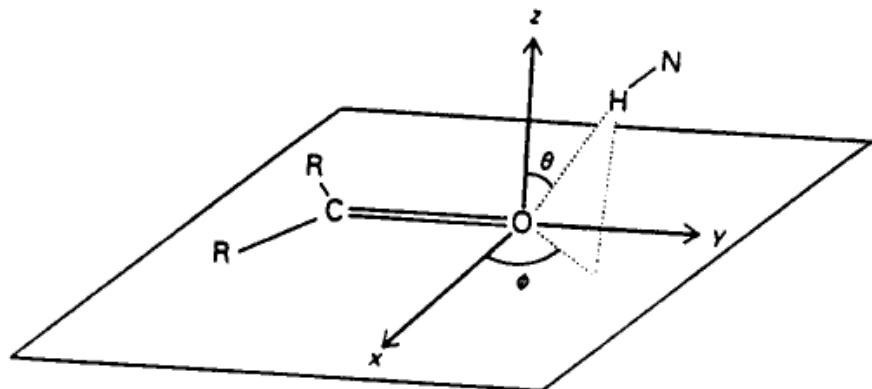


FIGURE 4.4

The geometries of  $\text{C}=\text{O} \cdots \text{H}-\text{N}$  hydrogen bonds observed in crystal structures of small molecules. The definitions of the angles  $\phi$  and  $\theta$  are illustrated at the top, and the relative frequencies of their observed values in intermolecular hydrogen bonds (R. Taylor et al., *J. Amer. Chem. Soc.* 105:5761–5766, 1983) are given by the contours. The angle  $\phi$  measures departures from linearity of the  $\text{C}=\text{O}$  bond and the H atom; the most frequently observed values are in the region of  $50^\circ$ – $60^\circ$ . The angle  $\theta$  measures the extent to which the H atom lies out of the plane defined by the R, C, and O atoms; the most commonly observed values are in the region of  $0^\circ$ – $7^\circ$ . The lone-pair electrons of the oxygen atom are believed to project at angles of  $\phi = 60^\circ$ ,  $\theta = 0^\circ$ . The spherical polar coordinate system used here gives a bias toward small values of  $\theta$  that could be corrected by plotting  $\sin \theta$ .

Table 4.7 Lengths of  $\text{H}-\text{N} \cdots \text{O}=\text{C}$  hydrogen bonds<sup>a</sup>

| Donor                      | Mean $\text{H} \cdots \text{O}$ Distance for Different Acceptors ( $\text{\AA}$ ) | Carboxyl <sup>b</sup> | Carboxylate <sup>c</sup> | Amide |
|----------------------------|---|-----------------------|--------------------------|-------|
| $\text{N}-\text{H}^d$      | $2.002 \pm 0.012$   | $1.928 \pm 0.012$     | $1.934 \pm 0.005$        |       |
| $\text{N}^+-\text{H}^e$    | $1.983 \pm 0.055$   | $1.869 \pm 0.028$     | $1.858 \pm 0.043$        |       |
| $\text{NH}_3^+$            | $1.916 \pm 0.041$   | $1.886 \pm 0.018$     | $1.988 \pm 0.075$        |       |
| $\text{R}-\text{NH}_2^+$   | $1.936 \pm 0.014$   | $1.841 \pm 0.008$     | $1.891 \pm 0.034$        |       |
| $\text{R}_2-\text{NH}_2^+$ | $1.887 \pm 0.047$   | $1.796 \pm 0.014$     | $1.793 \pm 0.070$        |       |
| $\text{R}_3-\text{NH}^+$   |   | $1.722 \pm 0.025$     | $1.845 \pm 0.014$        |       |

<sup>a</sup> The  $\text{N}-\text{H}$  distance is generally  $1.03 \text{ \AA}$ ; adding this value to the tabulated distances gives the distance between the N and O atoms.

<sup>b</sup>  $\text{C}=\text{O}$  oxygen atom of unionized carboxylic acids and esters.

<sup>c</sup> Oxygen atom of carboxyl anions ( $-\text{CO}_2^-$ ).

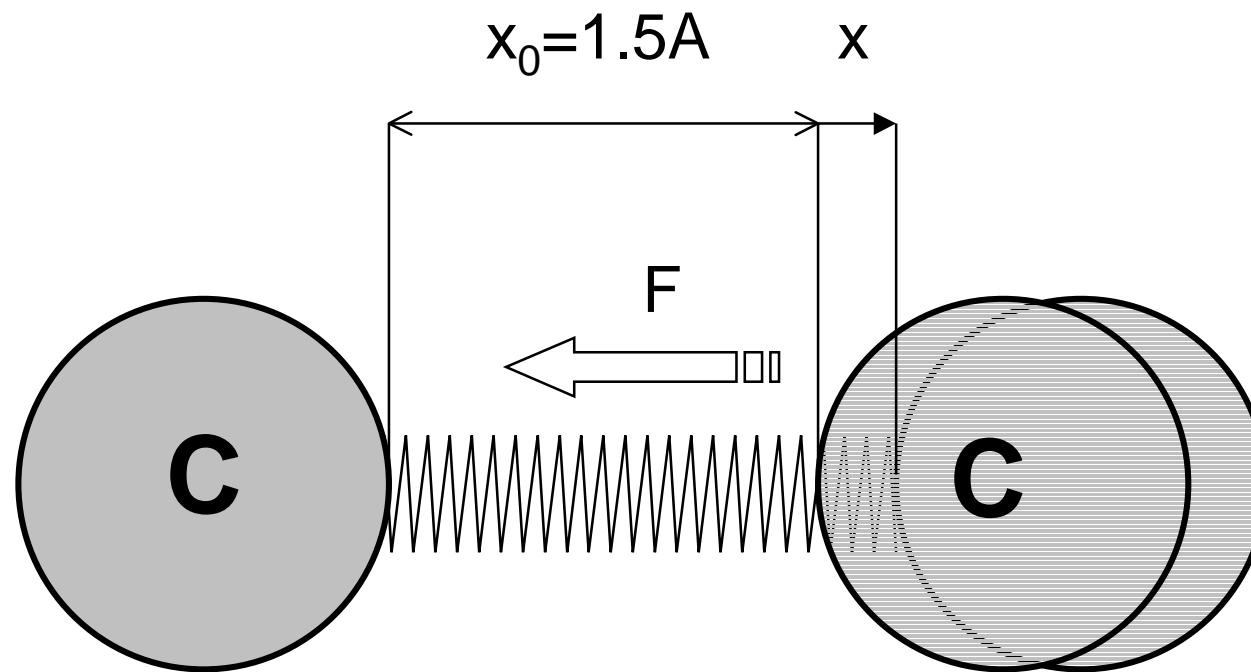
<sup>d</sup> Uncharged donor.

<sup>e</sup> Charged donor with trigonal geometry.

From R. Taylor and O. Kennard, *Acc. Chem. Res.* 17:320–326 (1984).

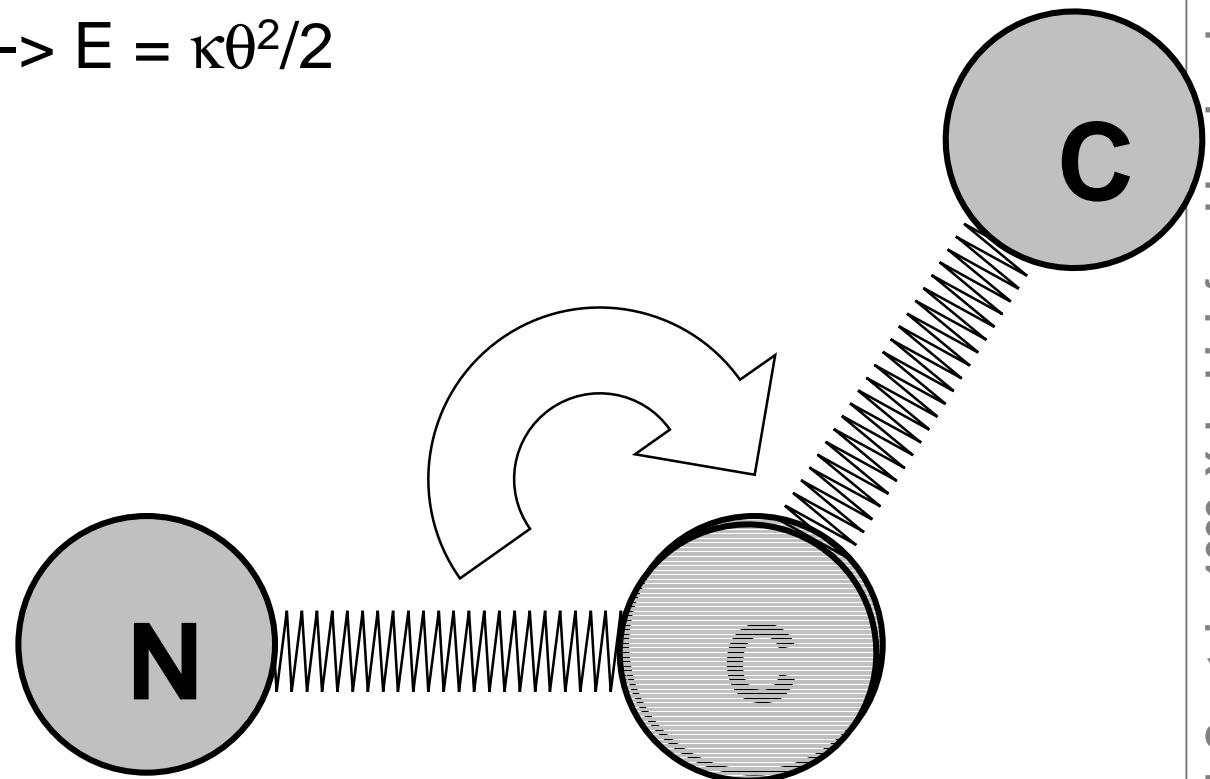
# Bond Length Springs

- $F = -kx \rightarrow E = kx^2/2$
- Freq from IR spectroscopy
  - ◊  $\omega = \sqrt{k/m}$ ,  $m$  = mass  $\Rightarrow$  spring const.  $k$
  - ◊  $k \sim 500 \text{ kcal/mole} \cdot \text{\AA}^2$  (stiff!),  
 $\omega$  corresponds to a period of 10 fs
- Bond length have 2-centers



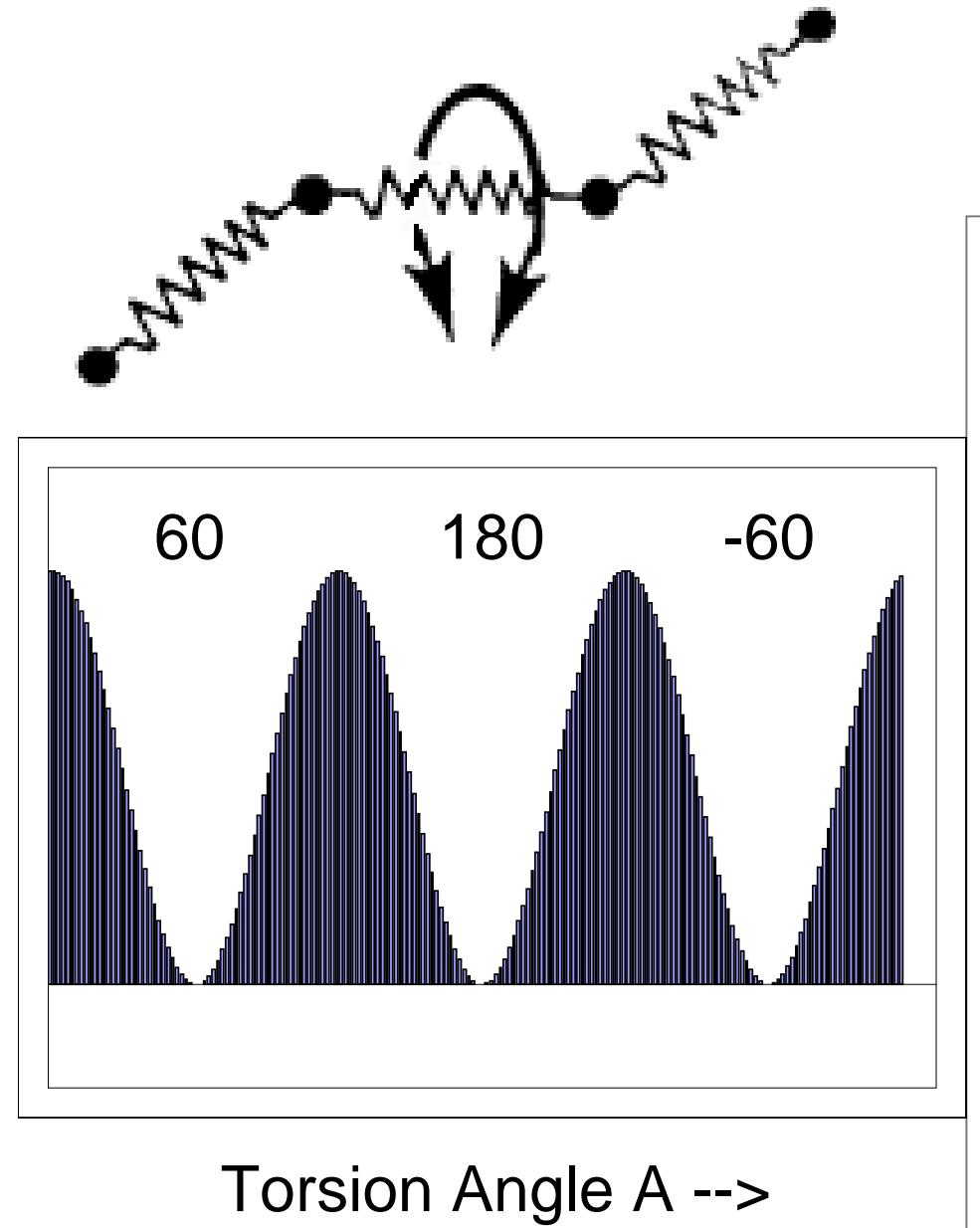
# Bond angle, More Springs

- torque =  $\tau = \kappa\theta \rightarrow E = \kappa\theta^2/2$
- 3-centers



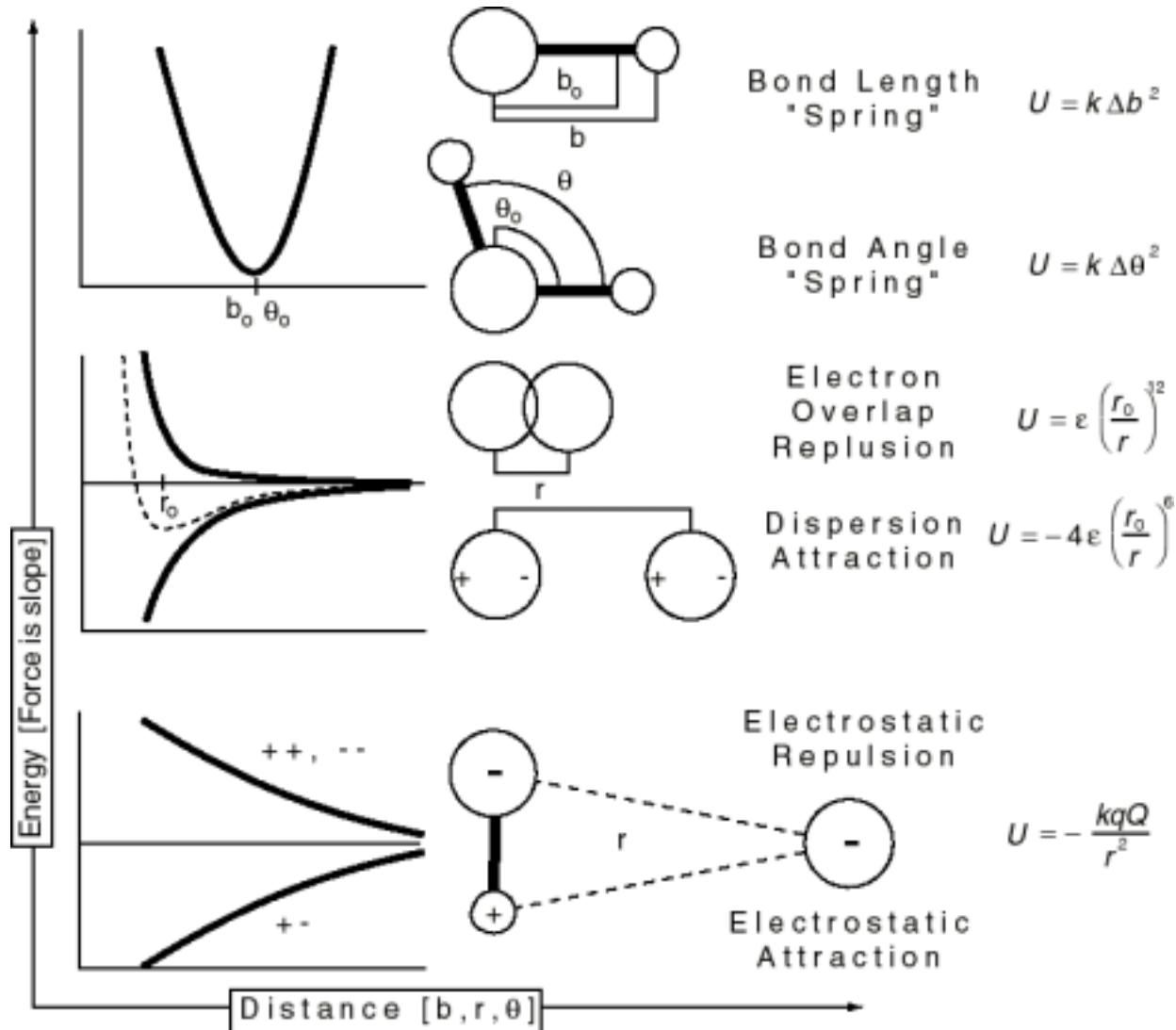
# Torsion angle

- 4-centers
- $U(A) = K(1 - \cos(nA + d))$ 
  - ◊  $\cos x = 1 - x^2/2 + \dots$ , so minima are quite spring like, but one can hoop between barriers
- $K \sim 2 \text{ kcal/mole}$



# Potential Functions

- Putting it all together
- Springs + Electrical Forces

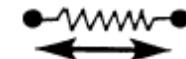


# Sum up to get total energy

- Each atom is a point mass ( $m$  and  $\mathbf{x}$ )
- Sometimes special pseudo-forces: torsions and improper torsions, H-bonds, symmetry.

$$E_{empirical} =$$

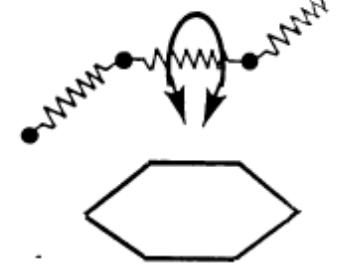
$$\sum_{bonds} k_o(b - b_o)^2$$



$$+ \sum_{angles} k_\Phi(\Phi - \Phi_o)^2$$



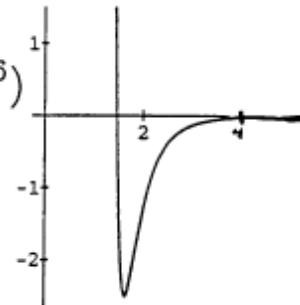
$$+ \sum_{dihedrals} k_\Psi \cos(n\Psi + \delta)$$



$$+ \sum_{chiral, planar\ centers} k_\omega(\omega - \omega_o)^2$$



$$+ \sum_{non-bonded} (Qr^{-1} + Ar^{-12} - Br^{-6})$$



$$+ \sum_{symmetry\ non-bonded} (Qr^{-1} + Ar^{-12} - Br^{-6})$$

# Energy Scale of Interactions

## THE SCALE OF INTERACTIONS

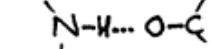
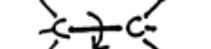
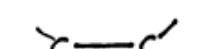
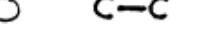
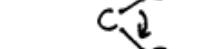
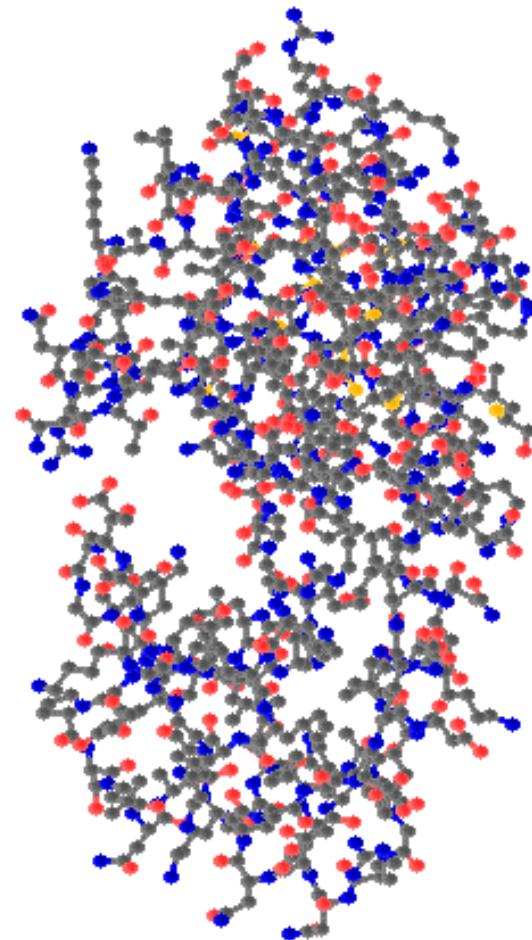
| Interaction                              | Energy (kcal/mole) |   |
|--|--------------------|---|
| van der Waals in water                   | -0.1               |    |
| van der Waals in vacuo                   | -0.3               |    |
| Hydrogen bond in water                   | -1.0               |    |
| Hydrogen bond in vacuo                   | -5.0               |    |
| Torsion barrier about single bond        | +3.0               |    |
| Torsion barrier about double bond        | +20                |    |
| Barrier to breaking a bond               | +100               |    |
| Energy to change a bond angle by 10°     | +2                 |   |
| Energy to stretch a bond length by 0.1 Å | +2.5               |  |
| Thermal energy at 300 K                  | 0.6                | $kT$  |

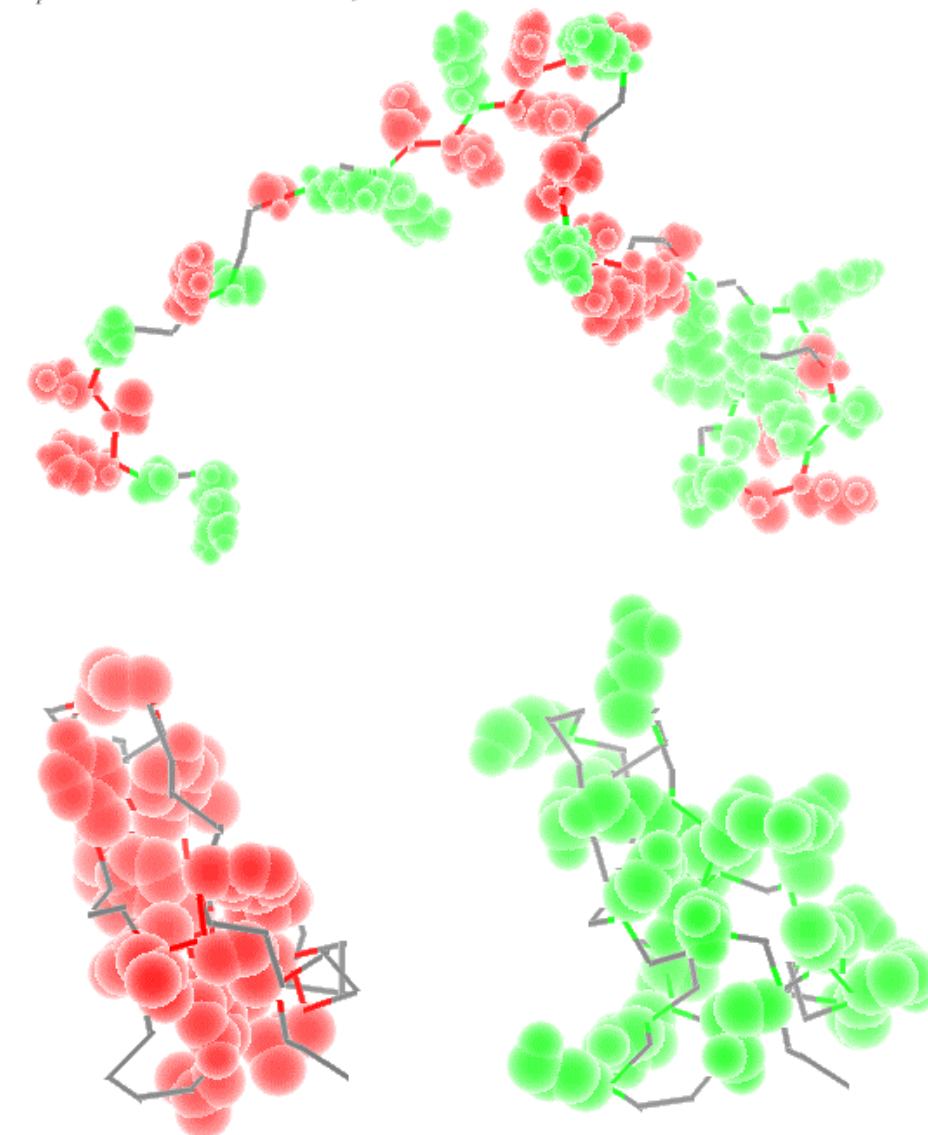
Illustration Credit: M Levitt

# Elaboration on the Basic Protein Model

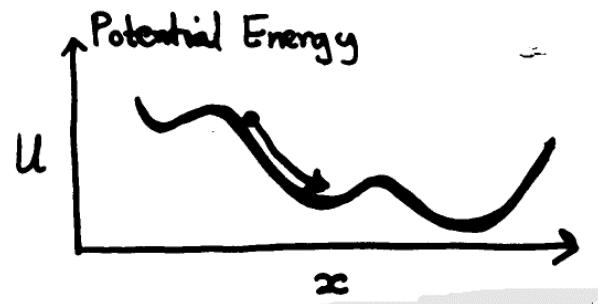
- Geometry
  - ◊ Start with X, Y, Z's (coordinates)
  - ◊ Derive Distance, Surface Area, Volume, Axes, Angle, &c
- Energetics
  - ◊ Add Q's and k's (Charges for electrical forces, Force Constants for springs)
  - ◊ Derive Potential Function  $U(x)$
- Dynamics
  - ◊ Add m's and t (mass and time)
  - ◊ Derive Dynamics  
( $v=dx/dt$ ,  $F = m dv/dt$ )



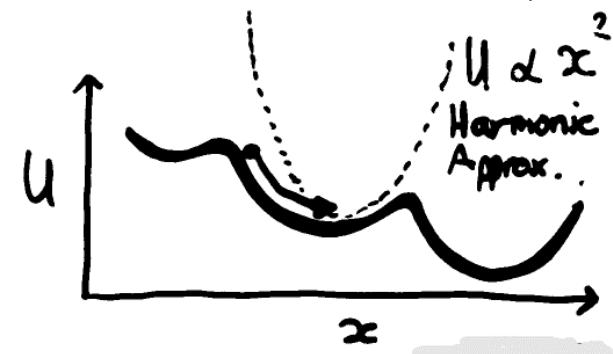
Goal:  
Model  
Proteins  
and  
Nucleic  
Acids  
as Real  
Physical  
Molecules



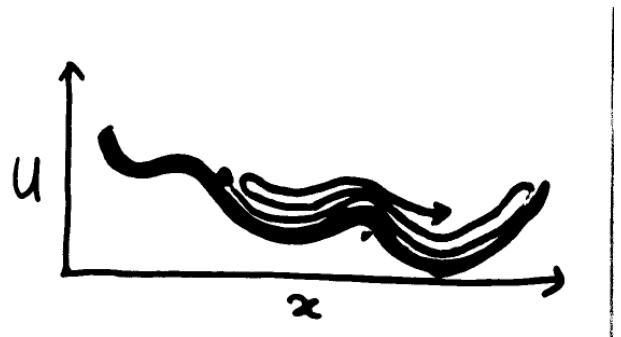
# Ways to Move Protein on its Energy Surface



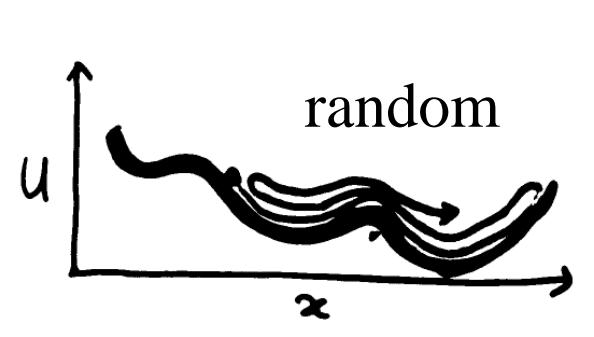
Minimization



Normal Mode Analysis (later?)



Molecular Dynamics (MD)

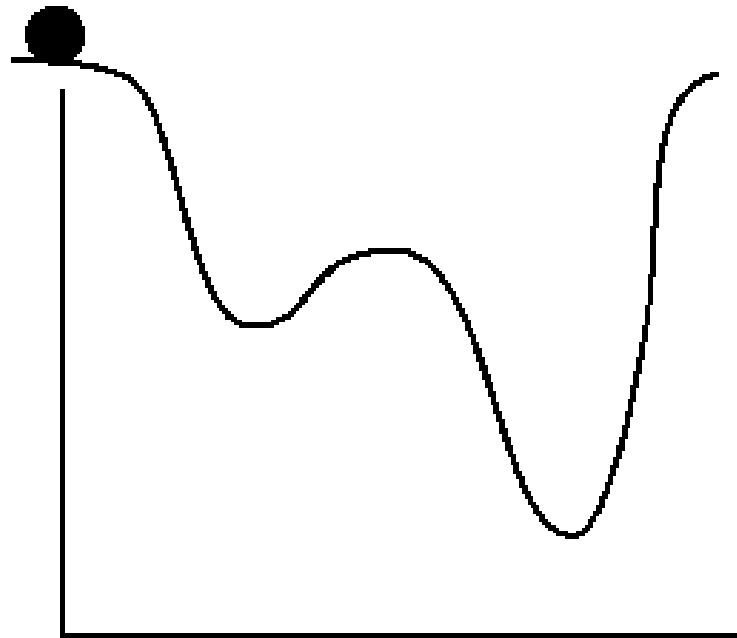


Monte Carlo (MC)

Illustration Credit: M Levitt

# Steepest Descent Minimization

- Particles on an “energy landscape.” Search for minimum energy configuration
  - ◊ Get stuck in local minima
- Steepest descent minimization
  - ◊ Follow gradient of energy straight downhill
  - ◊ i.e. Follow the force:  
**step** ~  $\mathbf{F} = -\nabla U$   
so  
$$\mathbf{x}(t) = \mathbf{x}(t-1) + a \mathbf{F}/|\mathbf{F}|$$



# Multi-dimensional Minimization

- In many dimensions, minimize along lines one at a time
- Ex:  $U = x^2 + 5y^2$  ,  $F = (2x, 10y)$

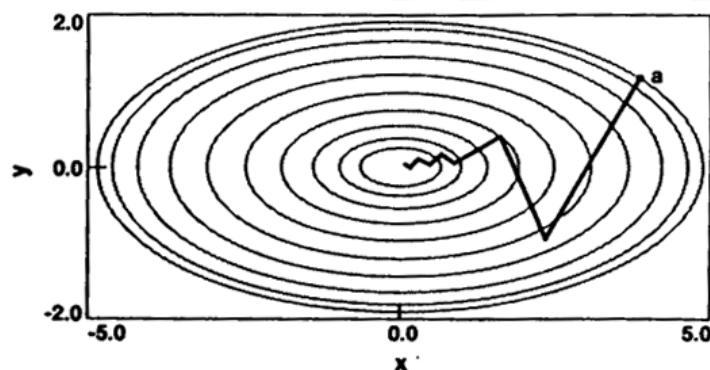


Figure 4-5. Minimization Path following a Steepest-Descents Path without Line Searches

The searching starts from point a and converges on the minimum in about 12 iterations. Although the number of iterations is slightly larger than in Figure 4-4, the total minimization is five times faster since, on average, each iteration used only 1.3 function evaluations. Note that, in most applications in molecular mechanics, the function evaluation is the most time-consuming portion of the calculation.

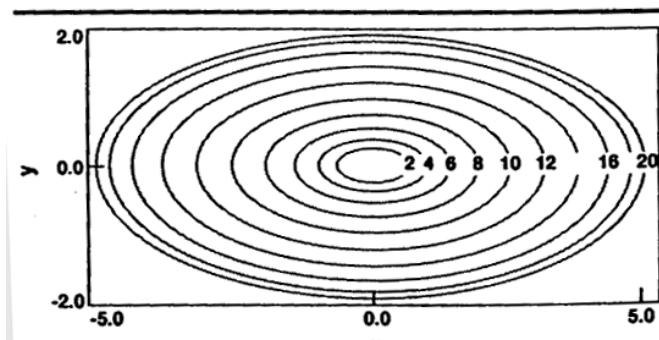


Figure 4-1. Energy Contour Surface of a Simple Function  
An energy contour surface for the function  $x^2 + 5y^2$ . Each contour represents an increase of two arbitrary energy units.

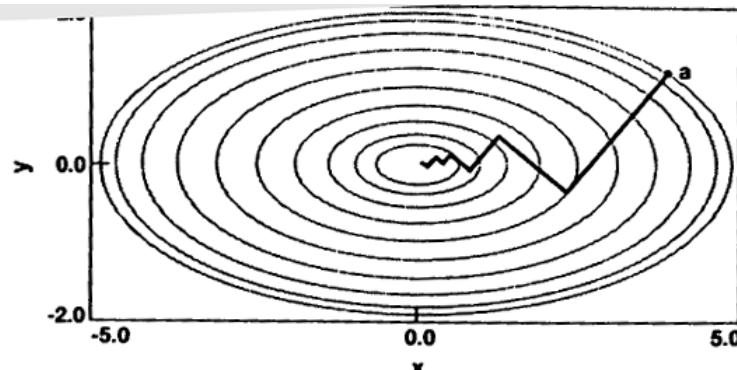


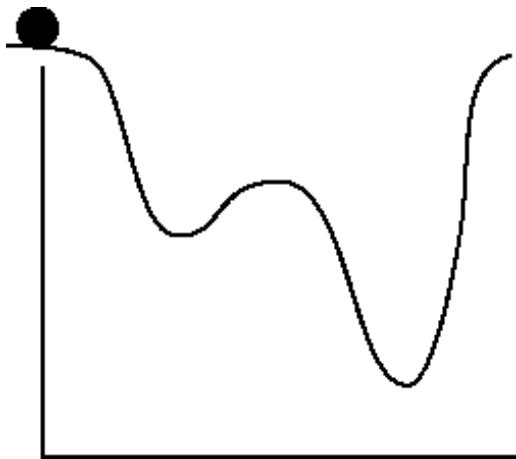
Figure 4-4. Minimization Path following a Steepest-Descents Path

When complete line searches starting from point a are used, the minimum is reached in about 12 iterations. Here, where a rigorous line search is carried out, approximately 8 function evaluations are needed for each line search using a quadratic interpolation scheme. Note how steepest descents consistently overshoots the best path to the minimum, resulting in an inefficient, oscillating trajectory.

Illustration Credit: Biosym, discover manual

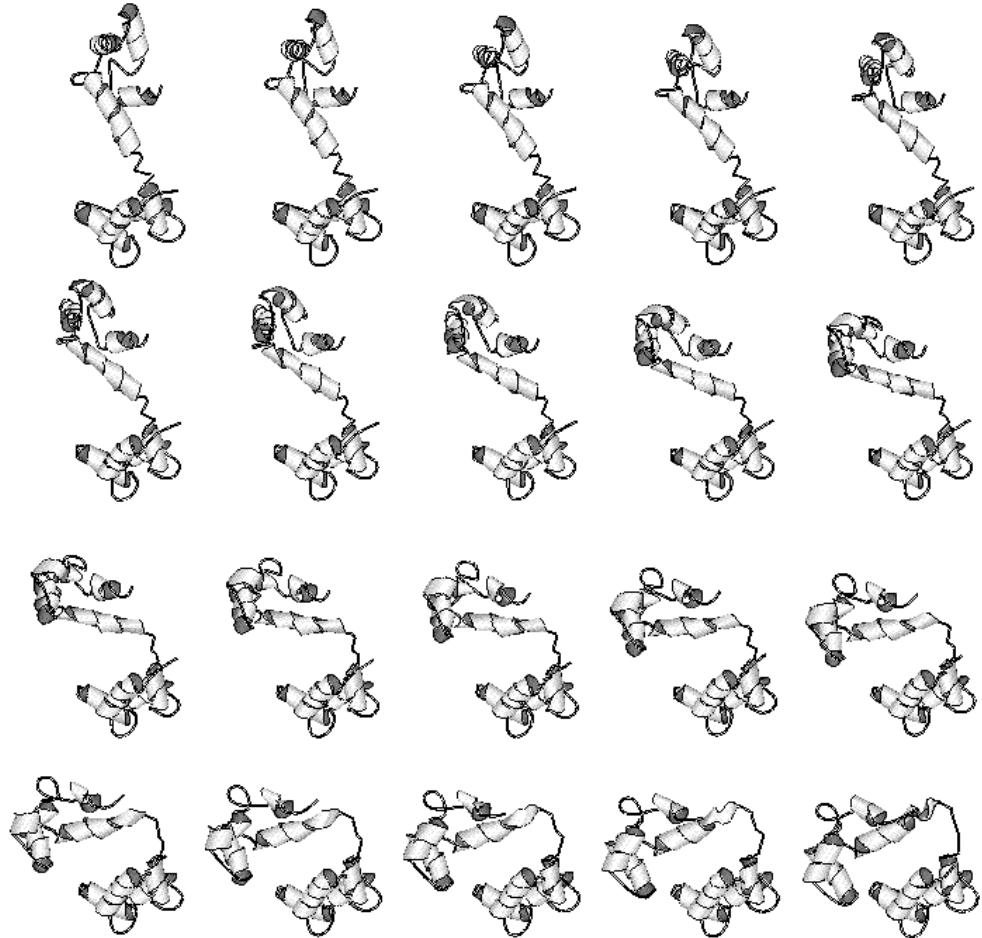
# Other Minimization Methods

- Simplex, grid search
  - ◊ no derivatives
- Conjugate gradient
  - step ~  $\mathbf{F}(t) - b\mathbf{F}(t-1)$**
  - ◊ partial 2nd derivative
- Newton-Raphson
  - ◊ using 2nd derivative, find minimum assuming it is parabolic
  - ◊  $V = ax^2 + bx + c$
  - ◊  $V' = 2ax + b$  &  $V'' = 2a$
  - ◊  $V' = 0 \rightarrow x^* = -b/2a$
- Problem is that get stuck in local minima
- Steepest descent, least clever but robust, slow at end
- Newton-Raphson faster but 2nd deriv. can be fooled by harmonic assumption
- Recipe: steepest descent 1st, then Newton-raph. (or conj. grad.)



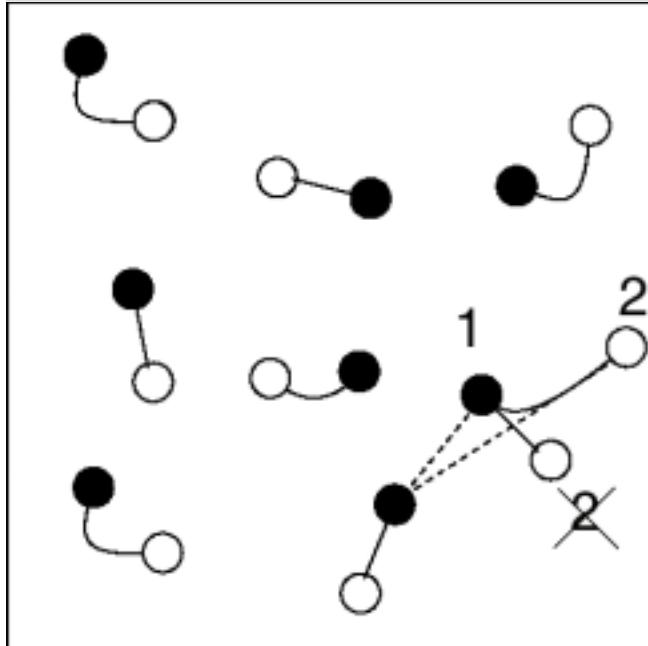
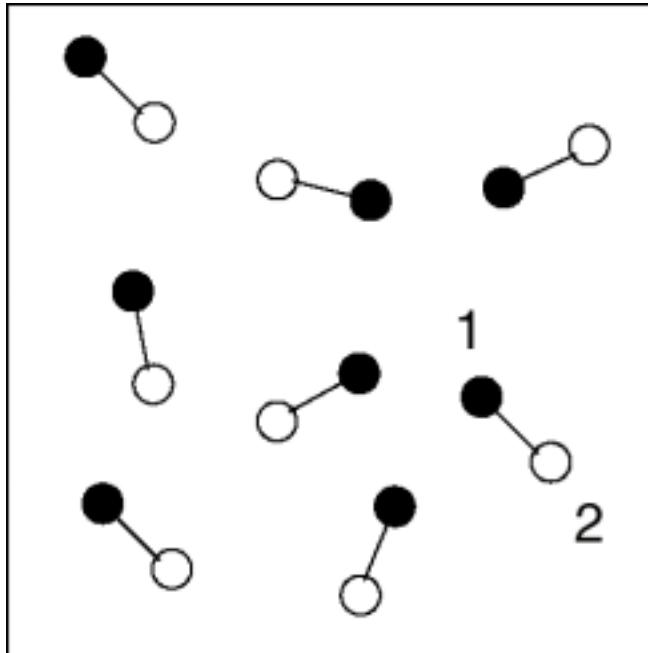
# Adiabatic mapping

- Interpolate then minimize
  - ◊ Gives apx. energy (H) landscape through a barrier
  - ◊ can sort of estimate transition rate  
$$\text{rate} = (kT/h) \exp(-dG/kT)$$
  - ◊ Used for ring flips, hinge motions



# Molecular Dynamics

- Give each atoms a velocity.
  - ◊ If no forces, new position of atom (at  $t + dt$ ) would be determined only by velocity  
 $\mathbf{x}(t+dt) = \mathbf{x}(t) + \mathbf{v} dt$
- Forces change the velocity, complicating things immensely
  - ◊  $\mathbf{F} = d\mathbf{p}/dt = m d\mathbf{v}/dt$



# Molecular Dynamics (cont)

- On computer make very small steps so force is nearly constant and velocity change can be calculated (uniform  $a$ )

$$\Delta \mathbf{v} = \frac{\mathbf{F}}{m} \Delta t$$

$$[\text{Avg. } \mathbf{v} \text{ over } \Delta t] = (\mathbf{v} + \Delta \mathbf{v}/2)$$

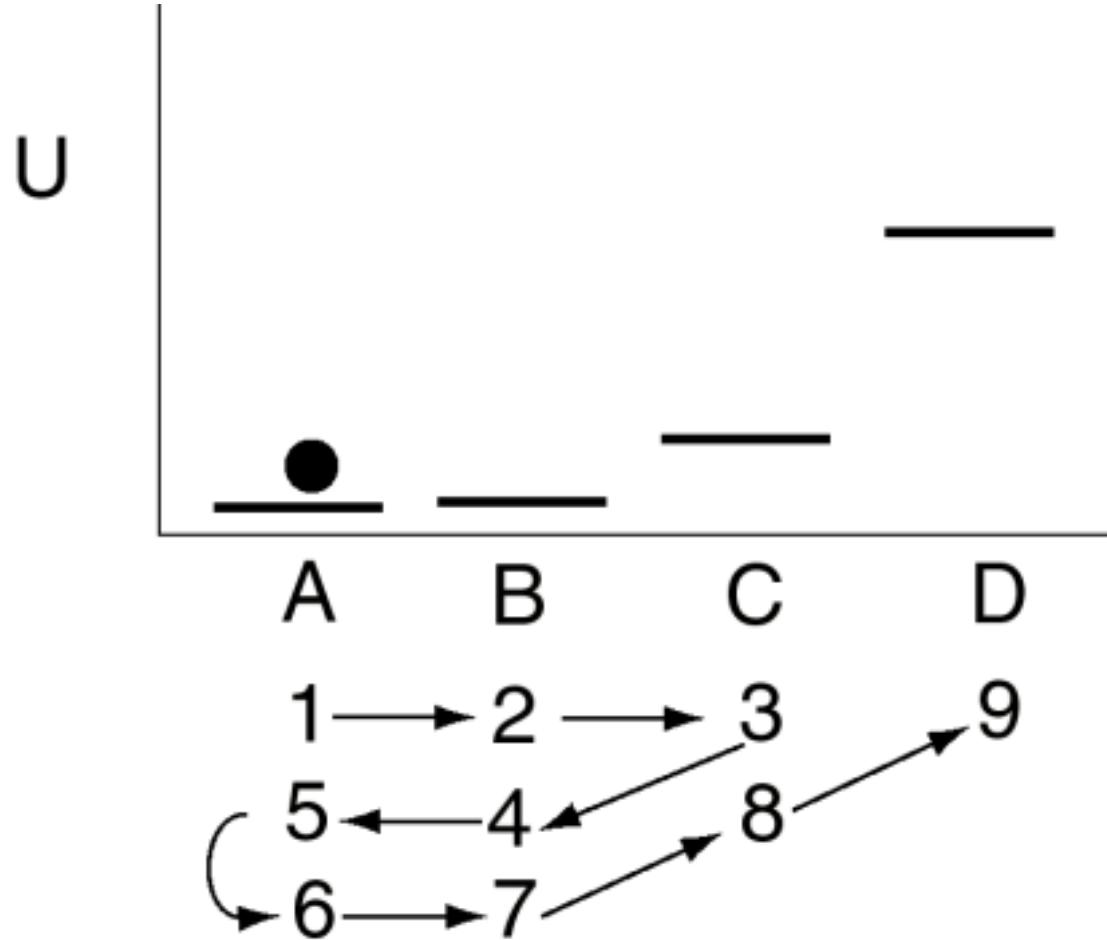
- Trivial to update positions:

$$\begin{aligned}\mathbf{x}(t + \Delta t) &= \mathbf{x}(t) + \left( \mathbf{v} + \frac{\Delta \mathbf{v}}{2} \right) \Delta t \\ &= \mathbf{x}(t) + \mathbf{v} \Delta t + \frac{\mathbf{F}}{2m} \Delta t^2\end{aligned}$$

- Step must be very small
  - ◊  $\Delta t \sim 1\text{fs}$   
(atom moves 1/500 of its diameter)
  - ◊ This is why you need fast computers
- Actual integration schemes slightly more complicated
  - ◊ Verlet (explicit half-step)
  - ◊ Beeman, Gear  
(higher order terms than acceleration)

# Phase Space Walk

- Trajectories of all the particles traverses space of all possible configuration and velocity states (phase space)
- Ergodic Assumption:  
Eventually, trajectory visits every state in phase space
- Boltzmann weighting:  
Throughout, trajectory samples states fairly in terms of system's energy levels
  - ◊ More time in low-U than high-U states
  - ◊ Probability of being in a state  $\sim \exp(-U/kT)$
- Consequently, statistics (average properties) over trajectory are thermodynamically correct



## Example Phase Space Walk

$$\langle X \rangle = 3X_A + 3X_B + 2X_A + 1X_D$$

$$\langle U \rangle = 6U_{AB} + 2U_A + 1U_D$$

# Monte Carlo

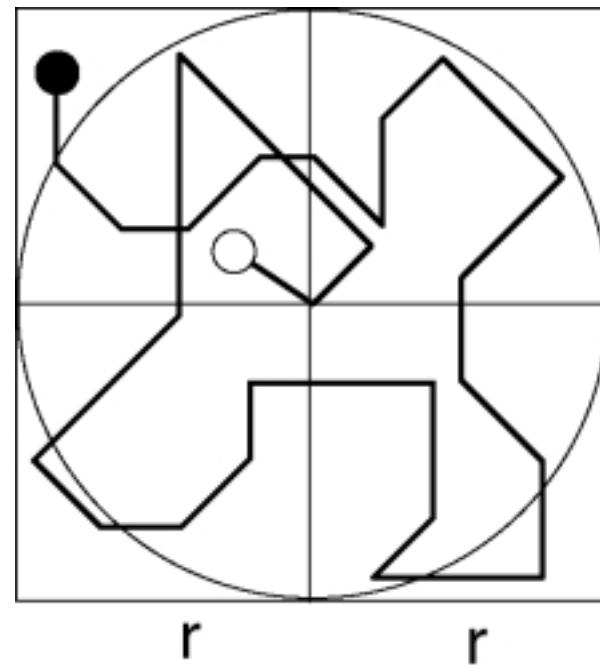
- Other ways than MD to sample states fairly and compute correctly weighted averages?  
Yes, using Monte Carlo calculations.
- Basic Idea:  
Move through states randomly, accepting or rejecting them so one gets a correct “Boltzmann weighting”
- Formalism:
  - ◊ System described by a probability distribution  $\rho(n)$  for it to be in each state n
  - ◊ Random (“Markov”) process  $\pi$  operates on the system and changes distribution amongst states to  $\pi\rho(n)$
  - ◊ At equilibrium original distribution and new distribution have to be same as Boltzmann distribution

$$\pi\rho(n) = \rho(n) = \frac{1}{Z} \exp\left(\frac{-U(n)}{kT}\right)$$

# Monte Carlo (cont)

- Metropolis Rule  
(for specifying  $\pi$ )
  - 1 Make a random move to a particle and calculate the energy change  $dU$
  - 2  $dU < 0 \rightarrow$  accept the move
  - 3 Otherwise, compute a random number  $R$  between 0 and 1:  
 $R < \sim \exp(-U/kT) \rightarrow$  accept the move  
otherwise  $\rightarrow$  reject the move

- “Fun” example of MC Integration
  - ◊ Particle in empty box of side  $2r$   
(energy of all states same)
  - ◊  $\pi = 6 \times [\text{Fraction of times particles is within } r \text{ of center}]$



## MC vs/+ MD

- MD usually used for proteins. Difficult to make moves with complicated chain.
- MC often used for liquids. Can be made into a very efficient sampler.
- Hybrid approaches (Brownian dynamics)
- Simulated Annealing. Heat simulation up to high T then gradually cool and minimize to find global minimum.

# Moving Molecules Rigidly

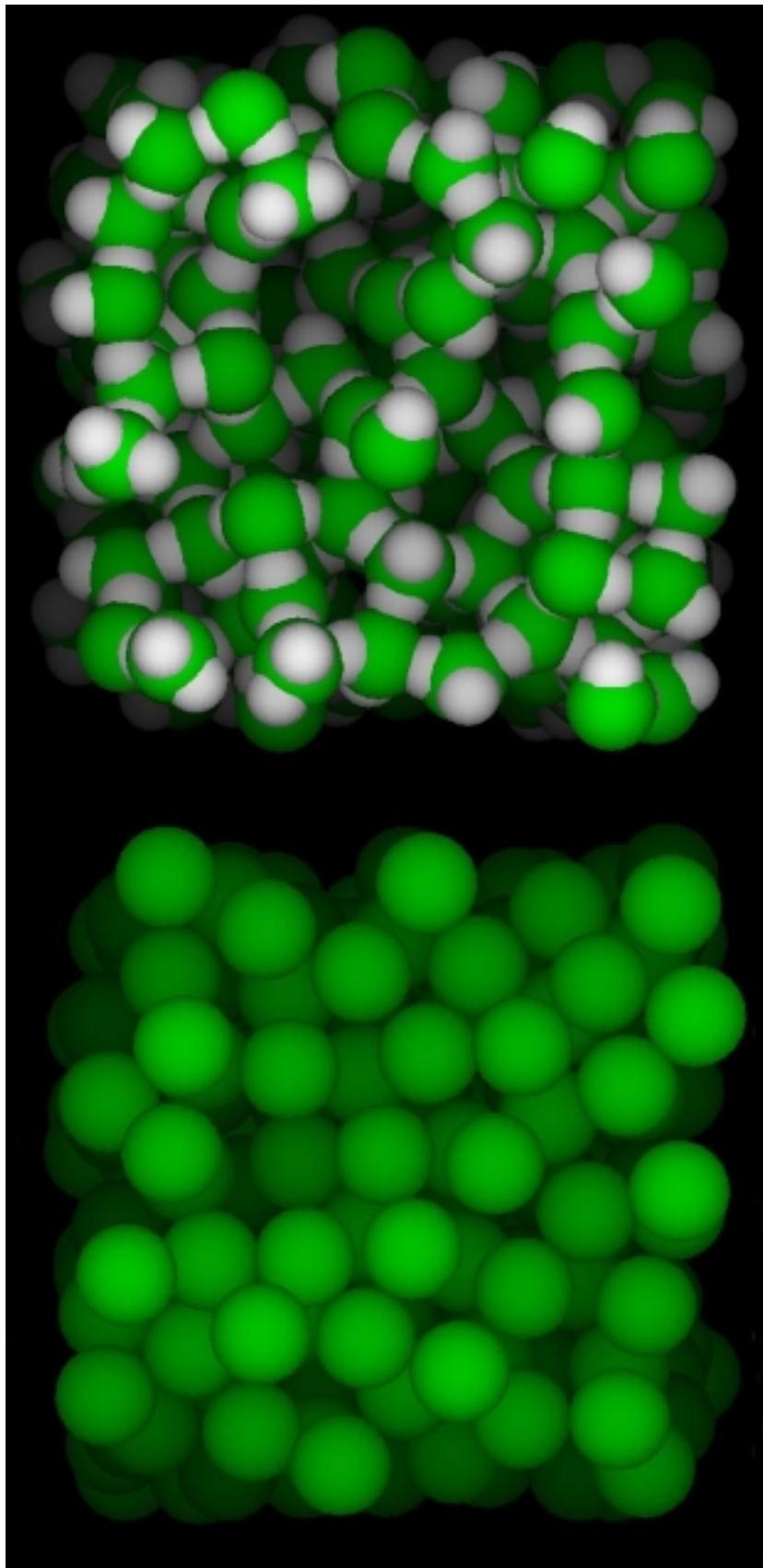
- $\mathbf{X}_i(t+1) = (x_i(t), y_i(t), z_i(t))$   
= coordinates of ith atom  
in the molecule at  
timestep t
- Rigid-body Translation of  
all i atoms
  - ◊ For each atom atom i do  
 $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{Finally, rotate by } \theta \text{ around z axis}} \underbrace{\begin{pmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{pmatrix}}_{\text{Second, rotate by } \phi \text{ around y axis}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{pmatrix}}_{\text{First, rotate by } \psi \text{ around x axis}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

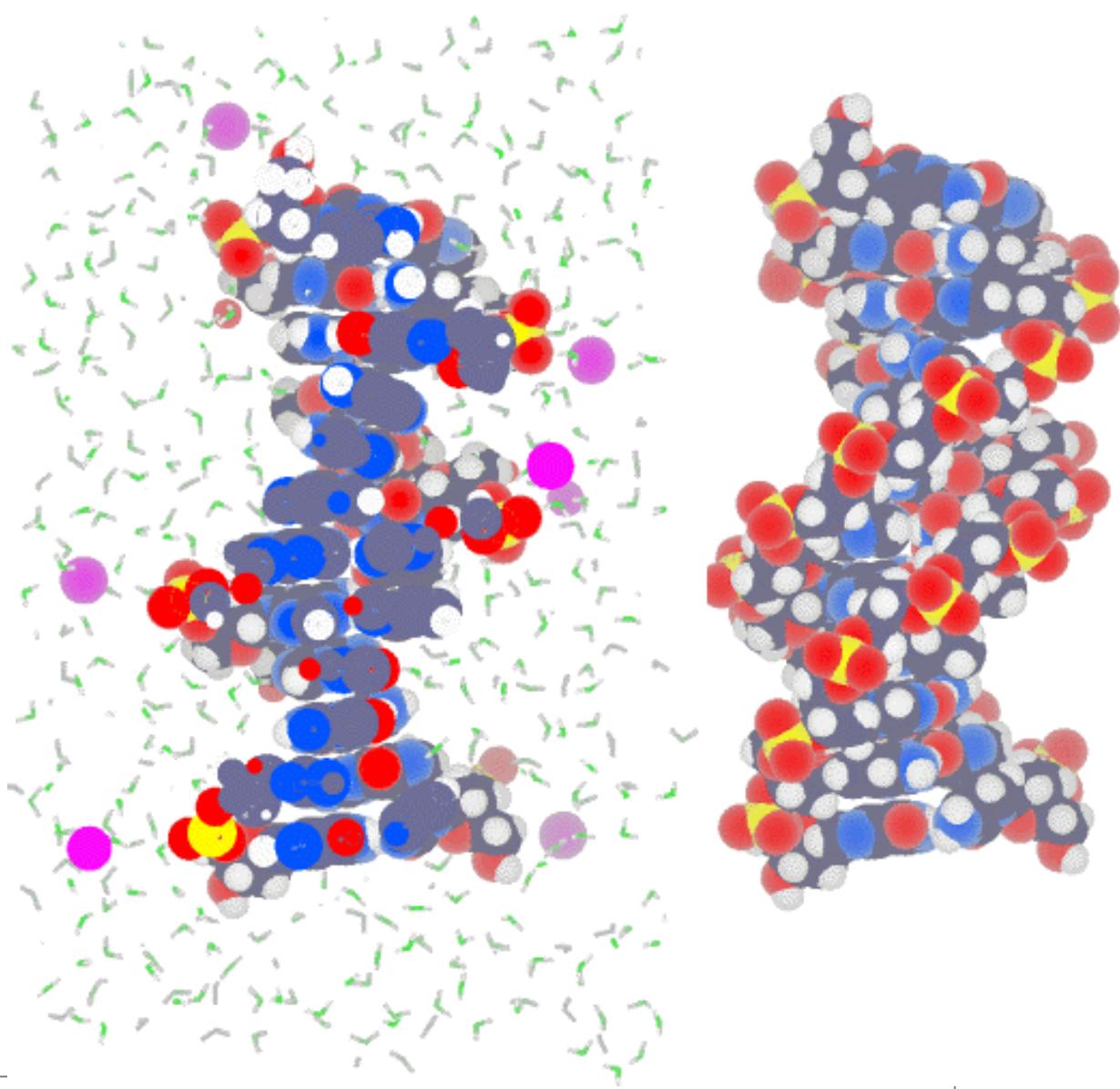
- Rigid-body Rotation of all i atoms
  - ◊ For each atom atom i do  
 $\mathbf{x}_i(t+1) = \mathbf{R}(\phi, \theta, \psi) \mathbf{x}_i(t)$
  - ◊ Effectively do a rotation around each axis (x, y, z)  
by angles  $\phi, \theta, \psi$  (see below)
  - ◊ Many conventions for doing this
    - **BELOW IS ONLY FOR MOTIVATION**
    - Consult Allen & Tildesley (1987) or Goldstein  
for the formulation of the rotation matrix  
using the usual conventions
  - ◊ How does one do a random rotation? Trickier  
than it seems

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

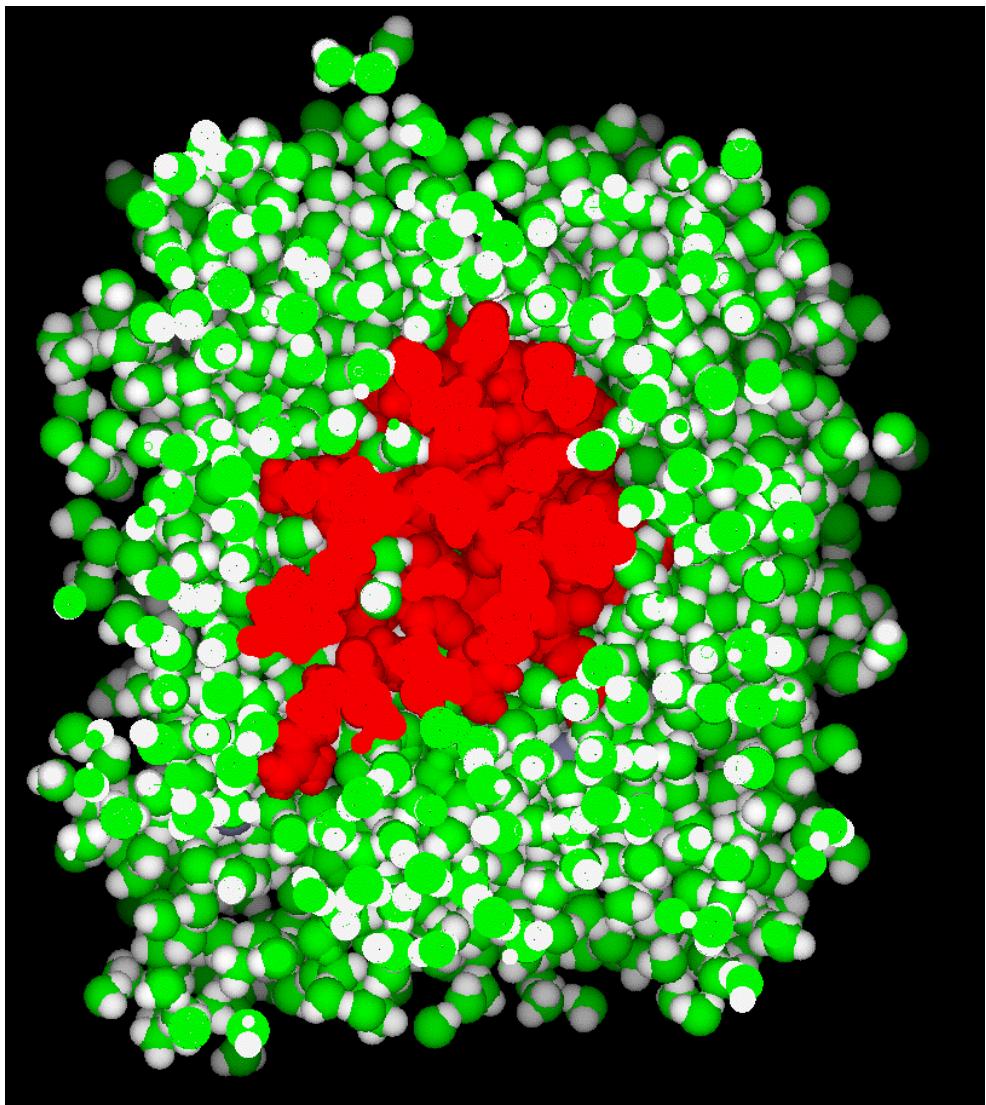
## Typical Systems: Water v. Argon



Typical  
Systems:  
DNA +  
Water



## Typical Systems: Protein + Water



# Practical Aspects: simulation cycle I

- Divide atoms into types (e.g. alpha carbon except for Gly, carbonyl oxygen)
- Initially
  - ◊ Associate each atom with a mass and a point charge
  - ◊ Give each atom an initial velocity
- Calculate Potential
- Calculating non-bonded interactions take up all the time
  - ◊ Electrostatics hardest since longest ranged
  - ◊ Neighbor lists

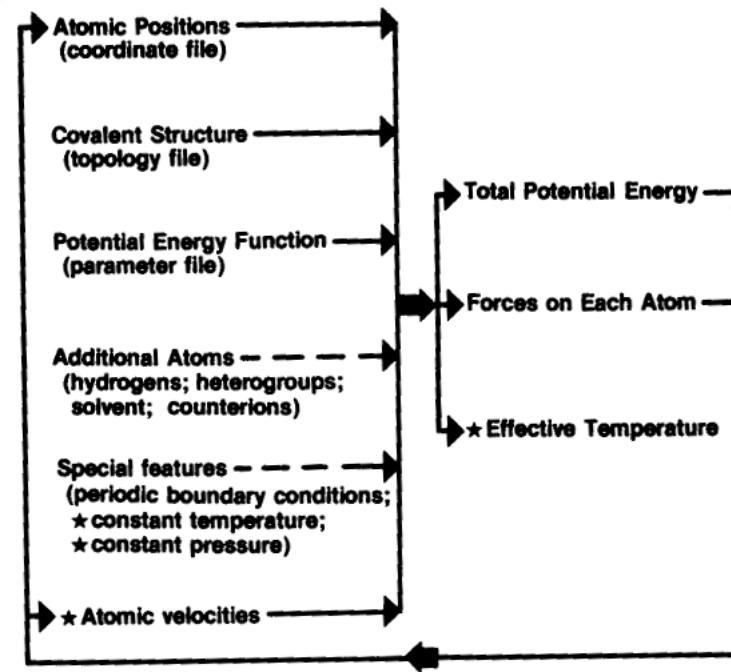


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of molecular dynamics represents a step in time.

Illustration Credit: McCammon & Harvey (1987)

# Practical Aspects: simulation cycle II

- Update Positions with MD equations, then recalculate potential and continue
- Momentum conservation
- Energy Conserved in NVE ensemble
- Hydrophobic interaction naturally arises from water behavior

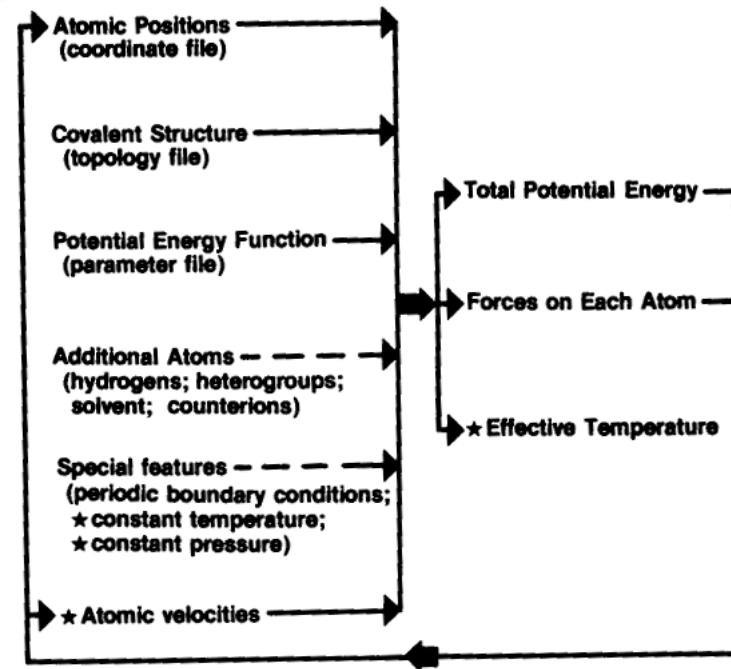


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of molecular dynamics represents a step in time.

Illustration Credit: McCammon & Harvey (1987)

# Sample Protein Parameters (toph19.pro)

```
REMARKS TOPH19.PRO ( protein topology )
REMARKS =====
REMARKS Charges and atom order modified for neutral GROUPs.
REMARKS Histidine charges set to Del Bene and Cohen sto-3g calculations.
REMARKS Amide charges set to match the experimental dipole moment.
REMARKS Default for HISTIDINES is the doubly protonated state

set echo=false end
!! for use with PARAM19 parameters ( no special hydrogen bonding potential )
!! donor and acceptor terms just for analysis

AUTOGENERATE ANGLES=TRUE END
{*=====*}

{* protein default masses *}
MASS H 1.00800! hydrogen which can h-bond to neutral atom
MASS HC 1.00800! = " = " = " = to charged atom
MASS HA 1.00800! aliphatic hydrogen
MASS CT 12.01100! aliphatic carbon
MASS C 12.01100! carbonyl carbon
MASS CH1E 13.01900! extended atom carbon with one hydrogen
MASS CH2E 14.02700! = " = " = two hydrogens
MASS CH3E 15.03500! = " = " = three hydrogens
MASS CR1E 13.01900! = " = " = in an aromatic ring with one H
MASS N 14.00670! peptide nitrogen with no hydrogens attached
MASS NR 14.00670! nitrogen in an aromatic ring with no hydrogens
MASS NP 14.00670! pyrole nitrogen
MASS NH1 14.00670! peptide nitrogen bound to one hydrogen
MASS NH2 14.00670! = " = " = two hydrogens
MASS NH3 14.00670! nitrogen bound to three hydrogens
MASS NC2 14.00670! charged guandinium nitrogen bound to two hydrogens
MASS O 15.99940! carbonyl oxygen
MASS OC 15.99940! carboxy oxygen
MASS OH1 15.99940! hydroxy oxygen
MASS S 32.06000! sulphur
MASS SH1E 33.06800! extended atom sulfur with one hydrogen

!some empirical rules for the following topologies:
!
```

# Sample Protein Parameters (toph19.pro)

```
. RESIDue ALA
GROUp
  ATOM N      TYPE=NH1      CHARge=-0.35      END
  ATOM H      TYPE=H       CHARge= 0.25      END
  ATOM CA     TYPE=CH1E     CHARge= 0.10      END
GROUp
  ATOM CB     TYPE=CH3E     CHARge= 0.00      END
GROUp
  ATOM C      TYPE=C       CHARge= 0.55      END  !#
  ATOM O      TYPE=O       CHARge=-0.55     END  !#
BOND N      CA
BOND CA     C
BOND C      O
BOND N      H
BOND CA     CB
IMPRoper   CA      N      C      CB      !tetrahedral CA
DONOr H      N
ACCEptor O  C
IC      N      C      *CA      CB      0.0000      0.00    120.00      0.00      0.0000
END { ALA }

!-----


RESIDue ARG
GROUp
  ATOM N      TYPE=NH1      CHARge=-0.35      END
  ATOM H      TYPE=H       CHARge= 0.25      END
  ATOM CA     TYPE=CH1E     CHARge= 0.10      END
GROUp
  ATOM CB     TYPE=CH2E     CHARge= 0.00      END
  ATOM CG     TYPE=CH2E     CHARge= 0.00      END
GROUp
  ATOM CD     TYPE=CH2E     CHARge= 0.10      END  !#
  ATOM NE     TYPE=NH1      CHARge=-0.40     END  !#
```

remark - parameter file PARAM19 -

bond C C 450.0 1.38! B. R. GELIN THESIS AMIDE AND DIPEPTIDES  
bond C CH1E 405.0 1.52! EXCEPT WHERE NOTED. CH1E,CH2E,CH3E, AND CT  
bond C CH2E 405.0 1.52! ALL TREATED THE SAME. UREY BRADLEY TERMS ADDED  
bond C CH3E 405.0 1.52  
bond C CR1E 450.0 1.38  
bond C CT 405.0 1.53  
bond C N 471.0 1.33  
bond C NC2 400.0 1.33! BOND LENGTH FROM PARMFIX9 FORCE K APPROXIMATE  
bond C NH1 471.0 1.33  
bond C NH2 471.0 1.33  
bond C NP 471.0 1.33  
bond C NR 471.0 1.33  
bond C O 580.0 1.23  
bond C OC 580.0 1.23! FORCE DECREASE AND LENGTH INCREASE FROM C O  
bond C OH1 450.0 1.38! FROM PARMFIX9 (NO VALUE IN GELIN THESIS)  
bond C OS 292.0 1.43! FROM DEP NORMAL MODE FIT  
bond CH1E CH1E 225.0 1.53  
bond CH1E CH2E 225.0 1.52  
bond CH1E CH3E 225.0 1.52  
bond CH1E N 422.0 1.45  
bond CH1E NH1 422.0 1.45  
bond CH1E NH2 422.0 1.45  
bond CH1E NH3 422.0 1.45  
bond CH1E OH1 400.0 1.42! FROM PARMFIX9 (NO VALUE IN GELIN THESIS)  
bond CH2E CH2E 225.0 1.52  
bond CH2E CH3E 225.0 1.54  
bond CH2E CR1E 250.0 1.45! FROM WARSHEL AND KARPLUS 1972 JACS 96:5612  
bond CH2E N 422.0 1.45  
bond CH2E NH1 422.0 1.45  
bond CH2E NH2 422.0 1.45  
bond CH2E NH3 422.0 1.45  
bond CH2E OH1 400.0 1.42  
bond CH2E S 450.0 1.81! FROM PARMFIX9  
bond CH2E SH1E 450.0 1.81  
  100.0 1.40

# Sample Protein Parameters (param19.pro)

# Sample Protein Parameters (param19.pro)

```
angle C   C   C      70.0 106.5! FROM B. R. GELIN THESIS WITH HARMONIC TERMS INCORPORATED. ATOMS WITH EXTENDED H COMPENSATED FOR LACK OF H ANGLES.
angle C   C   CH2E   65.0 126.5!
angle C   C   CH3E   65.0 126.5!
angle C   C   CR1E   70.0 122.5!
angle C   C   CT     70.0 126.5
angle C   C   HA     40.0 120.0! AMIDE PARAMETERS FIT BY LEAST SQUARES
angle C   C   NH1    65.0 109.0! TO N-METHYL ACETAMIDE VIBRATIONS
angle C   C   NP     65.0 112.5! MINIMIZATION OF N-METHYL ACETAMIDE.
angle C   C   NR     65.0 112.5
angle C   C   OH1    65.0 119.0
angle C   C   O      65.0 119.0 ! FOR NETROPSIN
angle CH1E C  N     20.0 117.5
angle CH1E C  NH1   20.0 117.5
angle CH1E C  O     85.0 121.5
angle CH1E C  OC    85.0 117.5
angle CH1E C  OH1   85.0 120.0
angle CH2E C  CR1E   70.0 121.5
angle CH2E C  N     20.0 117.5
angle CH2E C  NH1   20.0 117.5
angle CH2E C  NH2   20.0 117.5
angle CH2E C  NC2   20.0 117.5 ! FOR NETROPSIN
angle CH2E C  NR    60.0 116.0
angle CH2E C  O     85.0 121.6
angle CH2E C  OC    85.0 118.5
angle CH2E C  OH1   85.0 120.0
angle CH3E C  N     20.0 117.5
angle CH3E C  NH1   20.0 117.5
angle CH3E C  O     85.0 121.5
angle CR1E C  CR1E   65.0 120.5
angle CR1E C  NH1   65.0 110.5! USED ONLY IN HIS, NOT IT TRP
angle CR1E C  NP    65.0 122.5
angle CR1E C  NR    65.0 122.5
angle CR1E C  OH1   65.0 119.0
angle CT    C  N     20.0 117.5
angle CT    C  NH1   20.0 117.5
angle CT    C  NH2   20.0 117.5
angle CT    C  O     85.0 121.5
angle CT    C  OC    85.0 118.5
angle CT    C  OH1   85.0 120.0
angle HA    C  NH1   40.0 120.0
----- "A      40.0 120.0
```

```

!angle NR    FE    CM      5.0      180.0
!angle NR    FE    OM      5.0  180.0! JUST A GUESS FROM EXISTING FE CM DATA

```

# Sample Protein Parameters (param19.pro)

```

dihe CH1E C     N     CH1E 10.0      2      180.0! PRO ISOM. BARRIER 20 KCAL/MOL.
dihe CH2E C     N     CH1E 10.0      2      180.0
dihe CR1E C     C     CR1E 5.0       2      180.0! => TRP OOP. VIB 170CM 1
dihe CR1E C     C     C      2.5       2      180.0! SEE BEHLEN ET AL JCP 75:5685 81
dihe CR1E C     C     NH1   2.5       2      180.0
dihe X      C     CH1E X     0.0      3      0.0! FROM GELIN THESIS AMIDES
dihe X      C     CH2E X     0.0      3      0.0! USING A SINGLE
dihe X      C     CR1E X    10.0      2      180.0! DIHEDRAL PER BOND RATHER
dihe X      C     CT     X     0.0      3      0.0! THAN MULTIPLE TORSIONS
dihe X      C     N     X     8.2       2      180.0! ALKANE TORSION REDUCED TO
dihe X      C     NC2   X     8.2       2      180.0! 1.6 FROM 1.8 TO COINCIDE WITH
dihe X      C     NH1   X     8.2       2      180.0! THE EXPERIMENTAL BARRIER.
dihe X      C     NH2   X     8.2       2      180.0
dihe X      C     OH1   X     1.8       2      180.0
dihe X      C     OS    X     1.8       2      180.0 ! INFERRED FROM C-OH1
dihe X      CH1E CH1E X    1.6       3      0.0
dihe X      CH1E CH2E X    1.6       3      0.0
dihe X      CH1E N      X    0.3       3      0.0! FROM HAGLER ET AL TABULATION OF
dihe X      CH1E NH1   X    0.3       3      0.0! EXP. DATA AND 6 31G CALC.
dihe X      CH1E NH2   X    1.8       3      0.0! PROTONATED SECONDARY AMINE
dihe X      CH1E NH3   X    0.6       3      0.0! 1/PROTON SO 3 FOR THE BOND
dihe X      CH1E OH1   X    0.5       3      0.0! CHANGED TO ROUGHLY MEOH
dihe X      CH2E CH2E X    1.6       3      0.0
dihe X      CH2E N      X    0.3       3      0.0! SEE CH1E COMMENTS
dihe X      CH2E NH1   X    0.3       3      0.0
dihe X      CH2E NH2   X    0.6       3      0.0
dihe X      CH2E NH3   X    0.6       3      0.0
dihe X      CH2E OH1   X    0.5       3      0.0
dihe X      CH2E S      X    1.2       2      0.0
dihe X      CT     CT    X    1.6       3      0.0
dihe X      CT     N     X    0.3       3      0.0! SEE CH1E COMMENTS
dihe X      CT     NC2   X    0.3       3      0.0
dihe X      CT     NH1   X    0.3       3      0.0
dihe X      CT     NH2   X    0.6       3      0.0
dihe X      CT     NH3   X    0.6       3      0.0
dihe X      CT     OH1   X    0.5       3      0.0
dihe X      CT     S     X    1.2       2      0.0
!dihe X      FE     NR    X    0.05      4      0.0

```

# Sample Protein Parameters (param19.pro)

```

{* nonbonding parameter section *}
{* ===== *}
nbonds
atom cdie shift eps=1.0 e14fac=0.4 tolerance=0.5
cutnb=9.0 ctonnb=7.5 ctfnb=8.0
nbxmod=5 vswitch wmin 1.0
end

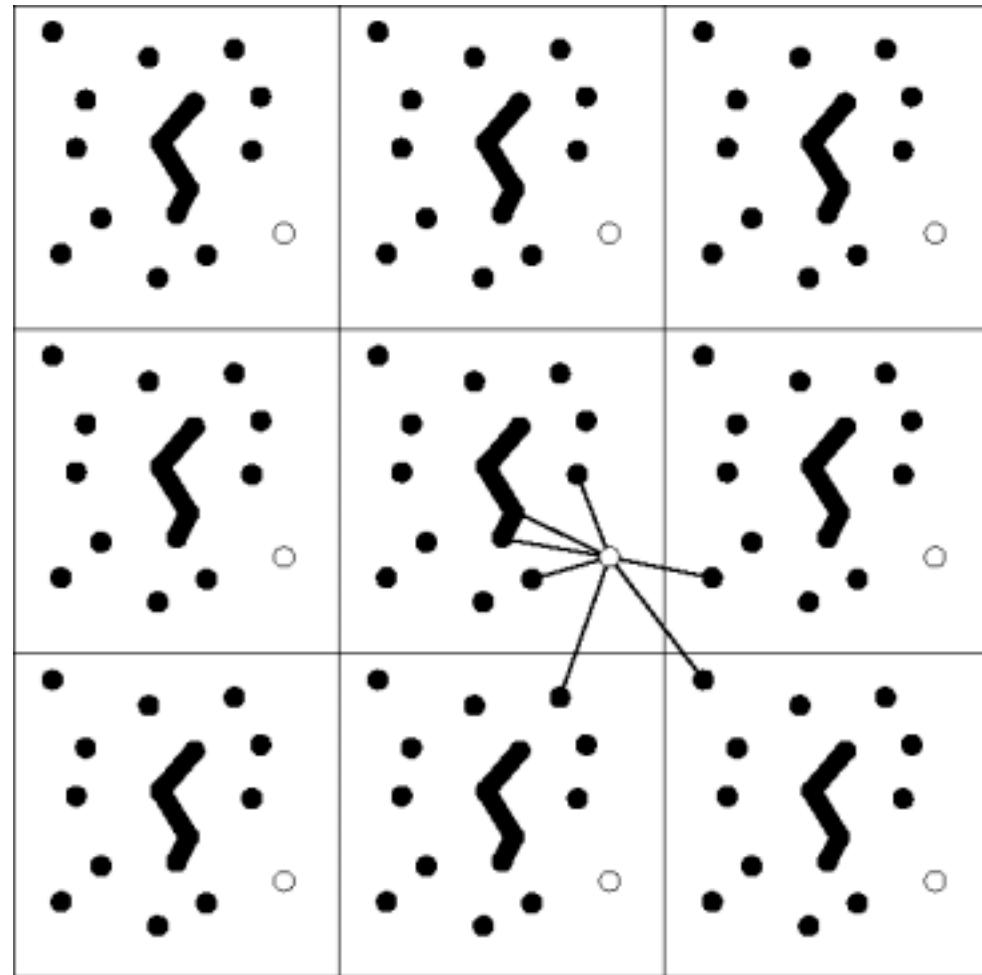
!
!           eps      sigma      eps(1:4) sigma(1:4)
!           (kcal/mol) (A)
!
-----  

NONBonded H    0.0498   1.4254    0.0498   1.4254
NONBonded HA   0.0450   2.6157    0.0450   2.6157 !- charged group.
NONBonded HC   0.0498   1.0691    0.0498   1.0691 ! Reduced vdw radius
!
NONBonded C    0.1200   3.7418    0.1000   3.3854 ! carbonyl carbon
NONBonded CH1E 0.0486   4.2140    0.1000   3.3854 ! \
NONBonded CH2E 0.1142   3.9823    0.1000   3.3854 ! extended carbons
NONBonded CH3E 0.1811   3.8576    0.1000   3.3854 ! /
!! NONBonded CM 0.0262   4.4367    0.1000   3.3854
NONBonded CR1E 0.1200   3.7418    0.1000   3.3854 ! ring carbons
!! NONBonded CT 0.0262   4.4367    0.1000   3.3854
!
NONBonded N    0.2384   2.8509    0.2384   2.8509
NONBonded NC2   0.2384   2.8509    0.2384   2.8509
NONBonded NH1   0.2384   2.8509    0.2384   2.8509
NONBonded NH2   0.2384   2.8509    0.2384   2.8509
NONBonded NH3   0.2384   2.8509    0.2384   2.8509
NONBonded NP    0.2384   2.8509    0.2384   2.8509
NONBonded NR    0.2384   2.8509    0.2384   2.8509
!
NONBonded O    0.1591   2.8509    0.1591   2.8509
NONBonded OC    0.6469   2.8509    0.6469   2.8509
NONBonded OH1   0.1591   2.8509    0.1591   2.8509
!! NONBonded OM 0.1591   2.8509    0.1591   2.8509
NONBonded OS    0.1591   2.8509    0.1591   2.8509
!
NONBonded S    0.0430   3.3676    0.0430   3.3676
NONBonded SH1E  0.0430   3.3676    0.0430   3.3676
!
!! NONBONDED FE        0.0000   1.1582    0.0000   1.1582
set a bostrue and

```

# Periodic Boundary Conditions

- Make simulation system seem larger than it is
- Ewald Summation for electrostatics (Fourier transform)



# Average over simulation

- Deceptive Instantaneous Snapshots  
(almost anything can happen)
- Simple thermodynamic averages
  - ◊ Average potential energy  $\langle U \rangle$
  - ◊  $T \sim \langle \text{Kinetic Energy} \rangle = \frac{1}{2} m \langle v^2 \rangle$
- Some quantities fixed, some fluctuate in different ensembles
  - ◊ NVE protein MD (“microcanonical”)
  - ◊ NVT liquid MC (“canonical”)
  - ◊ NPT more like the real world

---

| Motion                     | length time |           |
|----------------------------|-------------|-----------|
|                            | (Å)         | (fs)      |
| bond vibration             | 0.1         | 10        |
| water hindered rotation    | 0.5         | 1000      |
| surface sidechain rotation | 5           | $10^5$    |
| water diffusive motion     | 4           | $10^5$    |
| buried sidechain libration | 0.5         | $10^5$    |
| hinge bending of chain     | 3           | $10^6$    |
| buried sidechain rotation  | 5           | $10^{13}$ |
| allosteric transition      | 3           | $10^{13}$ |
| local denaturation         | 7           | $10^{14}$ |

Timescales

(From  
McCammon &  
Harvey,  
Eisenberg &  
Kauzmann)

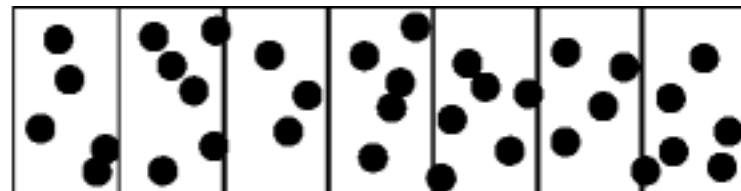
---

# D & RMS

- Diffusion constant
  - ◊ Measures average rate of increase in variance of position of the particles
  - ◊ Suitable for liquids, not really for proteins
- RMS more suitable to proteins
$$RMS(t) = \sqrt{\frac{\sum_{i=1}^N d_i(t)}{N}}$$
$$d_i(t) = \mathbf{R}(\mathbf{x}_i(t) - \mathbf{T}) - \mathbf{x}_i(0)$$
  - ◊  $d_i$  = Difference in position of protein atom at  $t$  from the initial position, after structures have been optimally rotated translated to minimize  $RMS(t)$
  - ◊ Solution of optimal rotation has been solved a number of ways (Kabsch, SVD)

$$D = \frac{\langle \Delta r^2 \rangle}{6\Delta t}$$

# Number Density



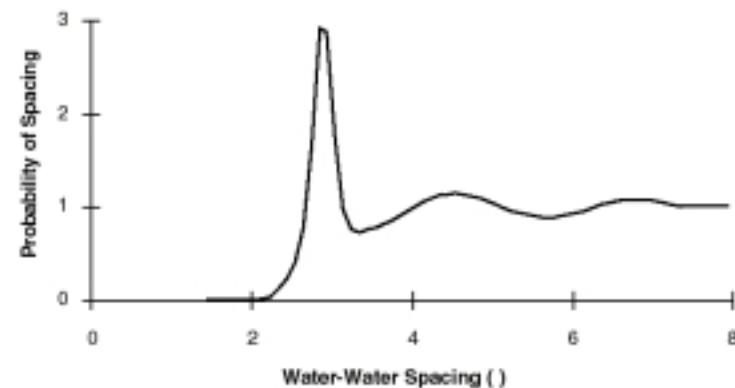
|  | Observed      | Expected      |
|--|---------------|---------------|
|  | $\frac{5}{5}$ | $\frac{5}{5}$ |
|  | $\frac{6}{5}$ | $\frac{5}{5}$ |
|  | $\frac{3}{5}$ | $\frac{5}{5}$ |
|  | $\frac{5}{5}$ | $\frac{5}{5}$ |
|  | $\frac{6}{5}$ | $\frac{5}{5}$ |
|  | $\frac{4}{5}$ | $\frac{6}{6}$ |

= Number of atoms per unit volume averaged over simulation divided by the number you expect to have in the same volume of an ideal “gas”

Spatially average over all directions gives

**1D RDF =**

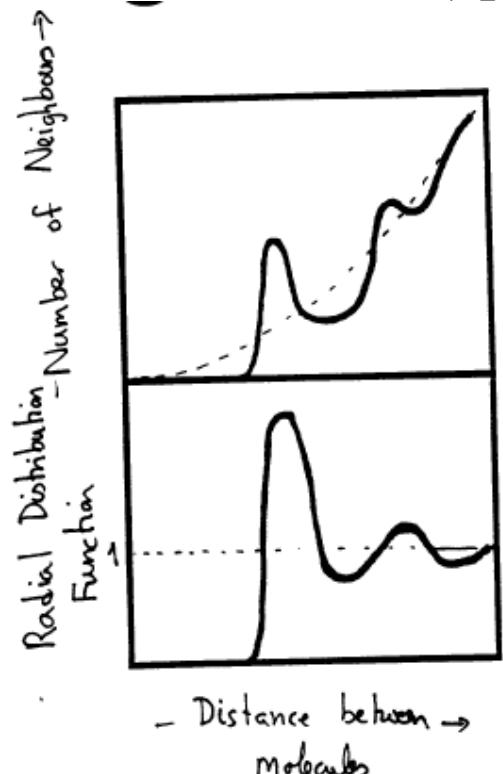
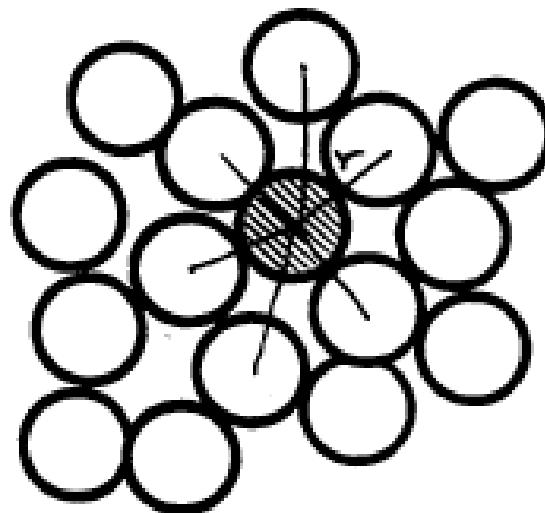
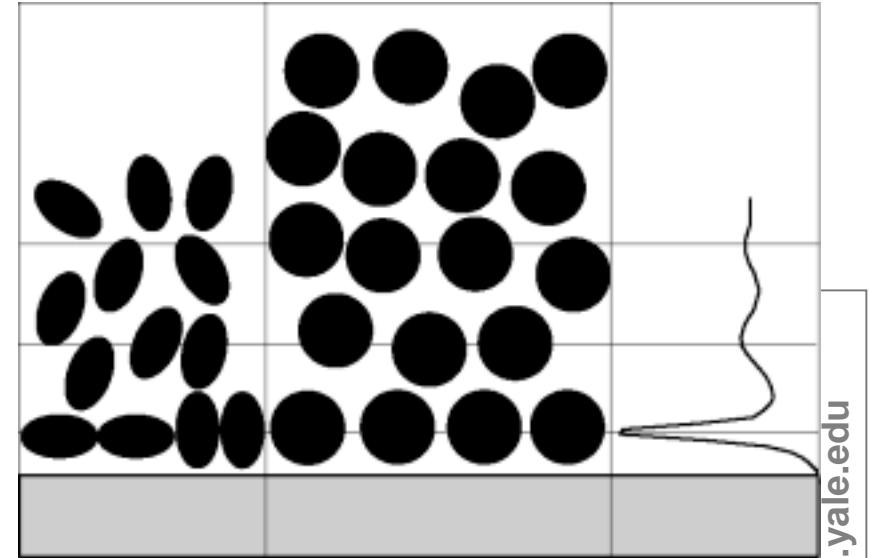
$$\frac{[\text{Avg. Num. Neighbors at } r]}{[\text{Expected Num. Neighbors at } r]}$$



“at  $r$ ” means contained in a thin shell of thickness  $dr$  and radius  $r$ .

# Number Density (cont)

- Advantages: Intuitive,  
Relates to scattering expts
- D/A: Not applicable to real  
proteins
  - ◊ 1D RDF not structural
  - ◊ 2D proj. only useful with "toy"  
systems
- Number densities  
measure spatial  
correlations, not packing
  - ◊ Low value does not imply  
cavities
  - ◊ Complicated by asymmetric  
molecules
  - ◊ How things pack and fit is  
property of instantaneous  
structure - not average



# Major Protein Simulation Packages

- AMBER
  - ◊ <http://www.amber.ucsf.edu/amber/amber.html>
  - ◊ <http://www.amber.ucsf.edu/amber/tutorial/index.html>
- CHARMM/XPLOR
  - ◊ <http://yuri.harvard.edu/charmm/charmm.html>
  - ◊ <http://atb.csb.yale.edu/xplor>
  - ◊ <http://uracil.cmc.uab.edu/Tutorials/default.html>
- ENCAD
- GROMOS
  - ◊ <http://rugmd0.chem.rug.nl/md.html>
  - ◊ “Advanced Crash Course on Electrostatics in Simulations” (!)  
(<http://rugmd0.chem.rug.nl/~berends/course.html>)

# Molecular Biophysics & Biochemistry

## 400a/700a (Advanced Biochemistry)

**Computational Aspects of:  
Simulation (Part II),  
Electrostatics (Part II),  
Water and Hydrophobicity**

Mark Gerstein

Classes on 11/12/98 & 10/17/98

Yale University

# The Handouts

- Notes
  - ◊ Coming on Tuesday!!!
  - ◊ Perhaps available on-line at <http://bioinfo.mbb.yale.edu/course>
- Presentation Paper
  - ◊ Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution *Science* **282**, 740-4.
    - <http://bioinfo.mbb.yale.edu/course/private-xxx/kollman-science-longsim.pdf>
    - <http://www.sciencemag.org/cgi/content/abstract/282/5389/740>
- Fun
  - ◊ Pollack, A. (1998). Drug Testers Turn to 'Virtual Patients' as Guinea Pigs. *New York Times*. Nov. 10
    - <http://www.nytimes.com/library/tech/98/11/biztech/articles/10health-virtual.html>
    - <http://bioinfo.mbb.yale.edu/course/private-xxx/pollack-nytimes-bioinfo.html>

# The Handouts II

- Review

- ◊ Sharp, K. (1999). Electrostatic Interactions in Proteins. In *International Tables for Crystallography*, International Union of Crystallography, Chester, UK.
- ◊ Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4**, 561-602.
- ◊ Gerstein, M. & Levitt, M. (1998). Simulating Water and the Molecules of Life. *Sci. Am.* **279**, 100-105.
  - <http://bioinfo.mbb.yale.edu/geometry/sciam>
- ◊ Franks, F. (1983). *Water*. The Royal Society of Chemistry, London. Pages 35-56.

- Homework Paper

- ◊ Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144-9.

# Outline

- Last Time

- ◊ Basic Forces
  - Electrostatics
  - Packing as VDW forces
  - Springs
- ◊ Minimization, Simulation

- Now

- ◊ Simulation, Part II: Analysis,  
What can be Calculated from Simulation?
- ◊ Electrostatics Revisited: the Poisson-Boltzmann Equation
- ◊ Water Simulation and Hydrophobicity
- ◊ Simplified Simulation

# Practical Aspects: simulation cycle I

- Divide atoms into types (e.g. alpha carbon except for Gly, carbonyl oxygen)
- Initially
  - ◊ Associate each atom with a mass and a point charge
  - ◊ Give each atom an initial velocity
- Calculate Potential
- Calculating non-bonded interactions take up all the time
  - ◊ Electrostatics hardest since longest ranged
  - ◊ Neighbor lists

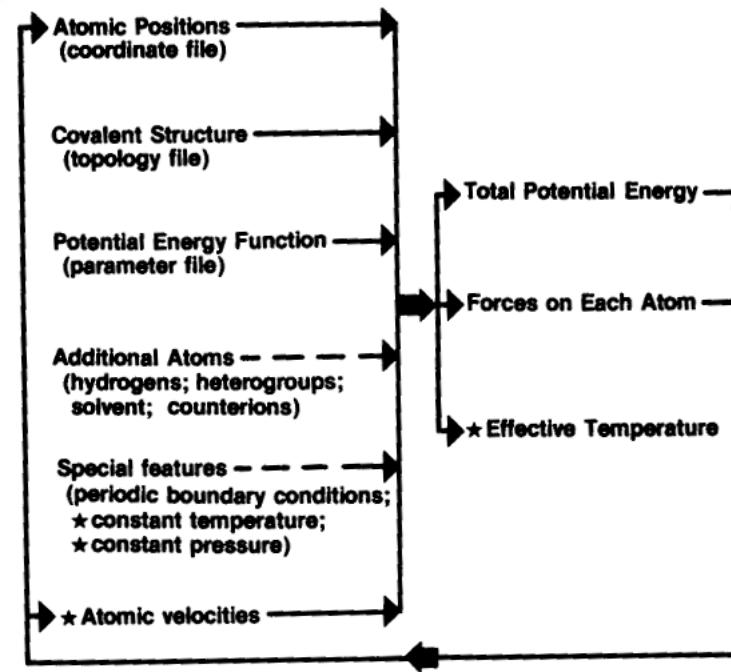


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of molecular dynamics represents a step in time.

Illustration Credit: McCammon & Harvey (1987)

# Practical Aspects: simulation cycle II

- Update Positions with MD equations, then recalculate potential and continue
- Momentum conservation
- Energy Conserved in NVE ensemble
- Hydrophobic interaction naturally arises from water behavior

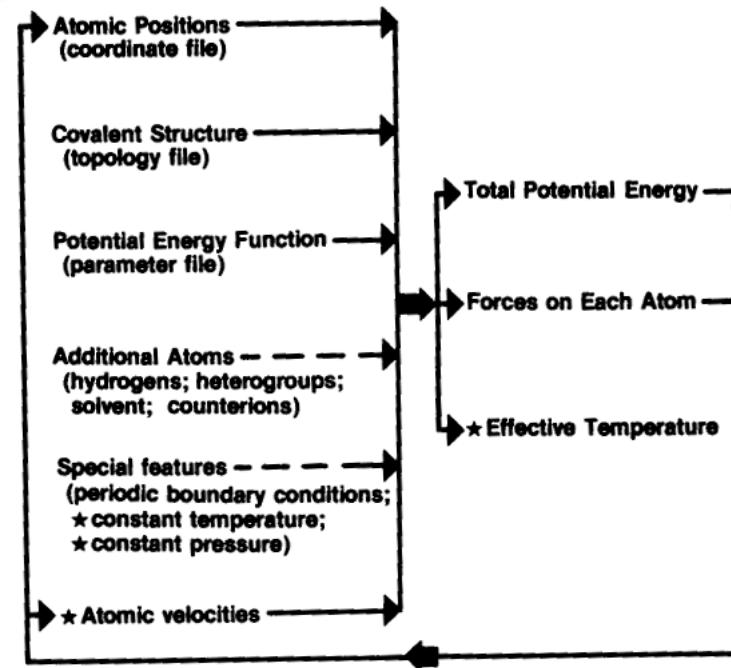


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of molecular dynamics represents a step in time.

Illustration Credit: McCammon & Harvey (1987)

# Major Protein Simulation Packages

- AMBER
  - ◊ <http://www.amber.ucsf.edu/amber/amber.html>
  - ◊ <http://www.amber.ucsf.edu/amber/tutorial/index.html>
- CHARMM/XPLOR
  - ◊ <http://yuri.harvard.edu/charmm/charmm.html>
  - ◊ <http://atb.csb.yale.edu/xplor>
  - ◊ <http://uracil.cmc.uab.edu/Tutorials/default.html>
- ENCAD
- GROMOS
  - ◊ <http://rugmd0.chem.rug.nl/md.html>
  - ◊ “Advanced Crash Course on Electrostatics in Simulations” (!)  
(<http://rugmd0.chem.rug.nl/~berends/course.html>)

# Moving Molecules Rigidly

- $\mathbf{X}_i(t+1) = (x_i(t), y_i(t), z_i(t))$   
= coordinates of ith atom  
in the molecule at  
timestep t
- Rigid-body Translation of  
all i atoms
  - ◊ For each atom atom i do  
 $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \underbrace{\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{Finally, rotate by } \theta \text{ around z axis}} \underbrace{\begin{pmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{pmatrix}}_{\text{Second, rotate by } \phi \text{ around y axis}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{pmatrix}}_{\text{First, rotate by } \psi \text{ around x axis}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

- Rigid-body Rotation of all i atoms
  - ◊ For each atom atom i do  
 $\mathbf{x}_i(t+1) = \mathbf{R}(\phi, \theta, \psi) \mathbf{x}_i(t)$
  - ◊ Effectively do a rotation around each axis (x, y, z)  
by angles  $\phi, \theta, \psi$  (see below)
  - ◊ Many conventions for doing this
    - **BELOW IS ONLY FOR MOTIVATION**
    - Consult Allen & Tildesley (1987) or Goldstein (1980) for the formulation of the rotation matrix using the usual conventions
  - ◊ How does one do a random rotation? Trickier than it seems

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

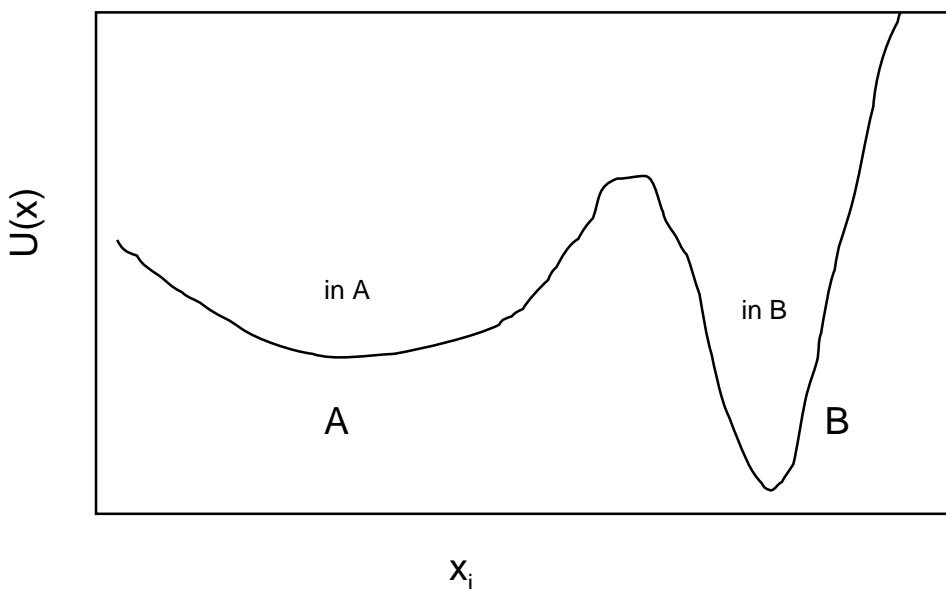
# Simulation, Part II: Analysis: What can be Calculated from Simulation?

# Average over simulation

- Deceptive Instantaneous Snapshots  
(almost anything can happen)
- Simple thermodynamic averages
  - ◊ Average potential energy  $\langle U \rangle$
  - ◊  $T \sim \langle \text{Kinetic Energy} \rangle = \frac{1}{2} m \langle v^2 \rangle$
- Some quantities fixed, some fluctuate in different ensembles
  - ◊ NVE protein MD (“microcanonical”)
  - ◊ NVT liquid MC (“canonical”)
  - ◊ NPT more like the real world

# Energy and Entropy

- Energy
  - ◊ At each point  $i$  (with coordinates  $x_i$ ) on the pot. energy surface there is a well-defined “energy”  $U(x_i)$
- Probability of occurrence
  - ◊  $P_i = \exp(-U_i/kT)/Q$
  - ◊ The boltzmann distribution
  - ◊  $Q = \text{Sum over all } P_i$ , to normalize probabilities to 1



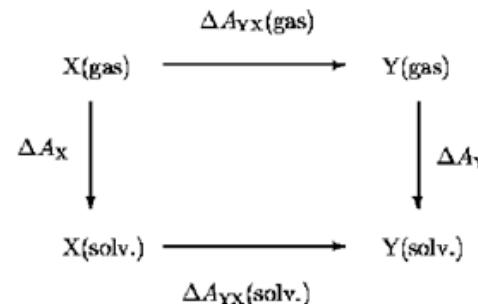
- Entropy
  - ◊  $S(A) = k \sum (P_i \ln P_i)$ , where the sum is over points  $i$  in A
- Free Energy
  - ◊  $G(A) = U(A) - TS(A)$
- Entropy and Free Energy are only defined for distinctly diff. “states” -- e.g. A (“unfolded”) and B (“folded”)
  - ◊ State B has a lower  $U$  and its minimum is more probable than State A
  - ◊ However, state A has a broader minimum that can be occupied in more ways
- Relative Prob
  - ◊  $P(A)/P(B) = \exp(-G(A)/kT)$

---
- $\exp (G(B)/kT)$

# Application of Simulation: Thermodynamic Cycles

## Molecular mutation

The difference of free energy of solvation  $\Delta\Delta\mu_{YX}$  between two solutes X and Y can be calculated by the following thermodynamic cycle:



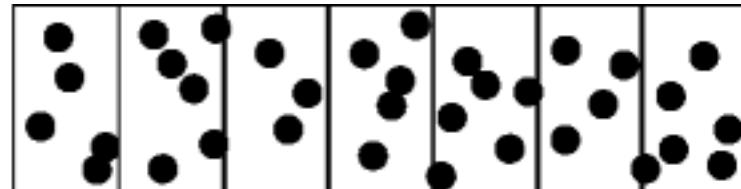
Text block adapted  
from on-line notes  
at Rutgers  
Chemistry

where  $\Delta\mu_X$  and  $\Delta\mu_Y$  are, respectively, the free energy of solvation of X and Y, and  $\Delta\mu_{YX}(\text{gas})$  and  $\Delta\mu_{YX}(\text{solv.})$  are the free energies of mutating X in Y in, respectively, in the gas phase and the solution phase.  
(Computational alchemy.)

The differences of free energies of solvation is

$$\Delta\Delta\mu_{YX} = \Delta\mu_Y - \Delta\mu_X = \Delta\mu_{YX}(\text{solv.}) - \Delta\mu_{YX}(\text{gas}) \quad (138)$$

# Number Density



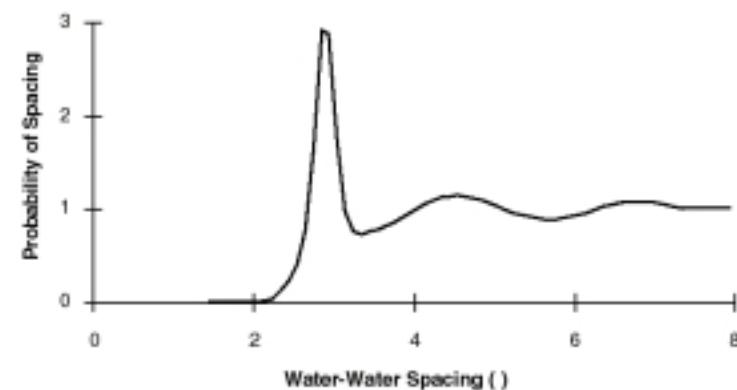
|  | Observed      | Expected      |
|--|---------------|---------------|
|  | $\frac{5}{5}$ | $\frac{5}{5}$ |
|  | $\frac{6}{5}$ | $\frac{5}{5}$ |
|  | $\frac{3}{5}$ | $\frac{5}{5}$ |
|  | $\frac{5}{5}$ | $\frac{5}{5}$ |
|  | $\frac{6}{5}$ | $\frac{5}{5}$ |
|  | $\frac{4}{5}$ | $\frac{6}{6}$ |

= Number of atoms per unit volume averaged over simulation divided by the number you expect to have in the same volume of an ideal “gas”

Spatially average over all directions gives

**1D RDF =**

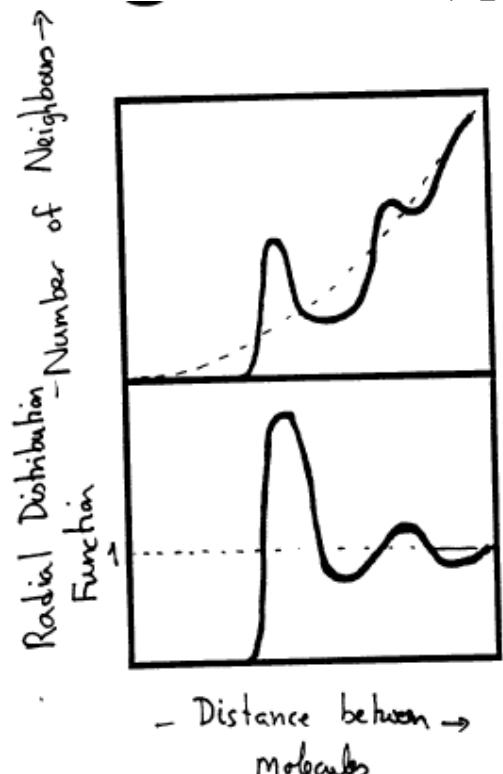
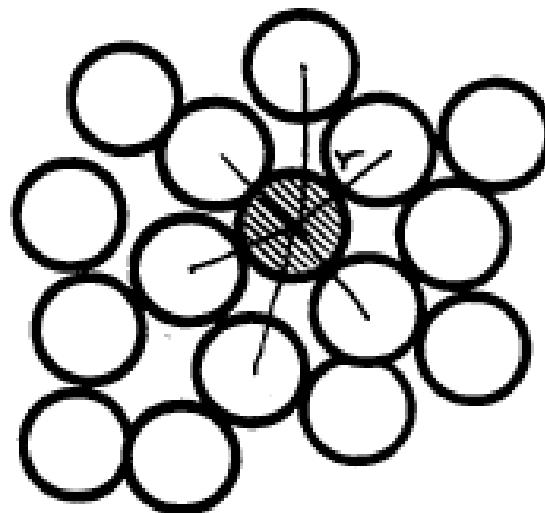
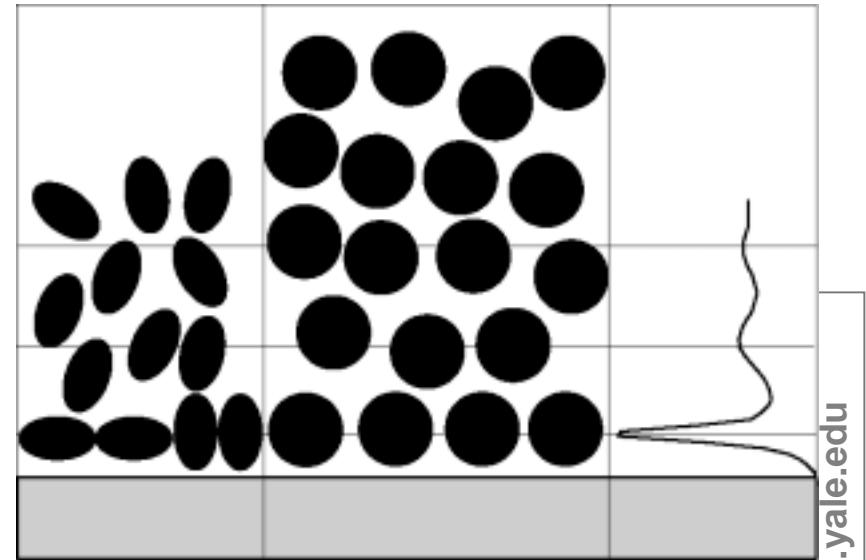
$$\frac{[\text{Avg. Num. Neighbors at } r]}{[\text{Expected Num. Neighbors at } r]}$$



“at  $r$ ” means contained in a thin shell of thickness  $dr$  and radius  $r$ .

# Number Density (cont)

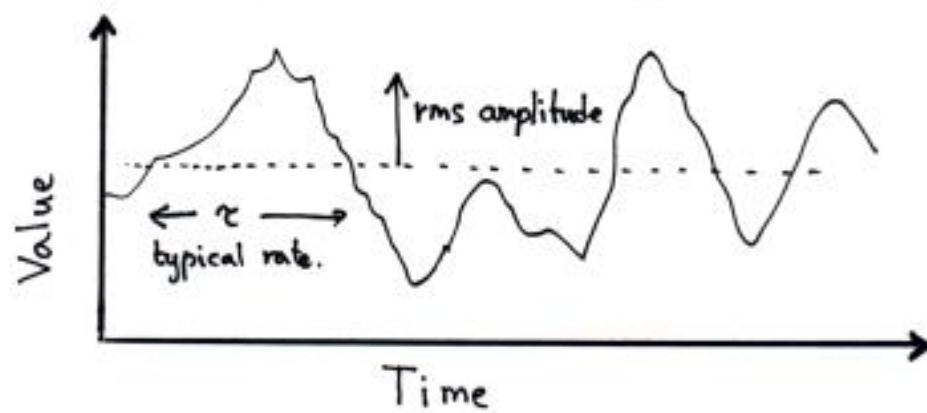
- Advantages: Intuitive,  
Relates to scattering expts
- D/A: Not applicable to real  
proteins
  - ◊ 1D RDF not structural
  - ◊ 2D proj. only useful with "toy"  
systems
- Number densities  
measure spatial  
correlations, not packing
  - ◊ Low value does not imply  
cavities
  - ◊ Complicated by asymmetric  
molecules
  - ◊ How things pack and fit is  
property of instantaneous  
structure - not average



# Measurement of Dynamic Quantities I

- The time-course of a relevant variable is characterized by
  - (1) Amplitude (or magnitude), usually characterized by an RMS value
$$R = \sqrt{ \langle (a(t) - \langle a(t) \rangle)^2 \rangle }$$
$$R = \sqrt{ \langle a(t)^2 \rangle - 2\langle a(t) \rangle \langle a(t) \rangle + \langle a(t) \rangle^2 }$$
$$R = \sqrt{ \langle a(t)^2 \rangle - \langle a(t) \rangle^2 }$$
    - similar to SD
    - fluctuation
- Relevant variables include bond length, solvent molecule position, H-bond angle, torsion angle

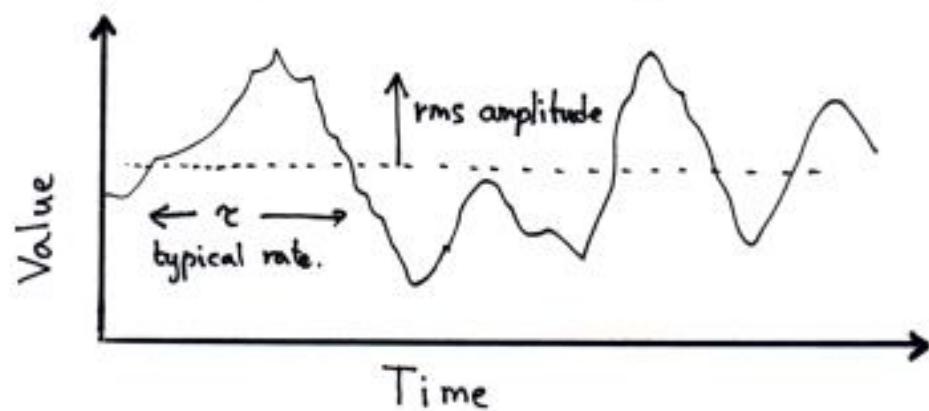
Illustration from M Levitt,  
Stanford University



# Measurement of Dynamic Quantities II

- The time-course of a relevant variable is characterized by
- (2) Rate or time-constant
- ◊ Time Correlation function
  - ◊  $C_A(t) = \langle A(s)A(t+s) \rangle = \langle A(0)A(t) \rangle$  [ averaging over all s ]
  - ◊ Correlation usually exponentially decays with time t
  - ◊ decay constant is given by the integral of  $C(t)$  from  $t=0$  to  $t=\infty$
- Relevant variables include bond length, solvent molecule position, H-bond angle, torsion angle

Illustration from M Levitt,  
Stanford University



# D & RMS

- Diffusion constant
  - ◊ Measures average rate of increase in variance of position of the particles
  - ◊ Suitable for liquids, not really for proteins
- RMS more suitable to proteins
$$RMS(t) = \sqrt{\frac{\sum_{i=1}^N d_i(t)}{N}}$$
$$d_i(t) = \mathbf{R}(\mathbf{x}_i(t) - \mathbf{T}) - \mathbf{x}_i(0)$$
  - ◊  $d_i$  = Difference in position of protein atom at  $t$  from the initial position, after structures have been optimally rotated translated to minimize  $RMS(t)$
  - ◊ Solution of optimal rotation has been solved a number of ways (Kabsch, SVD)

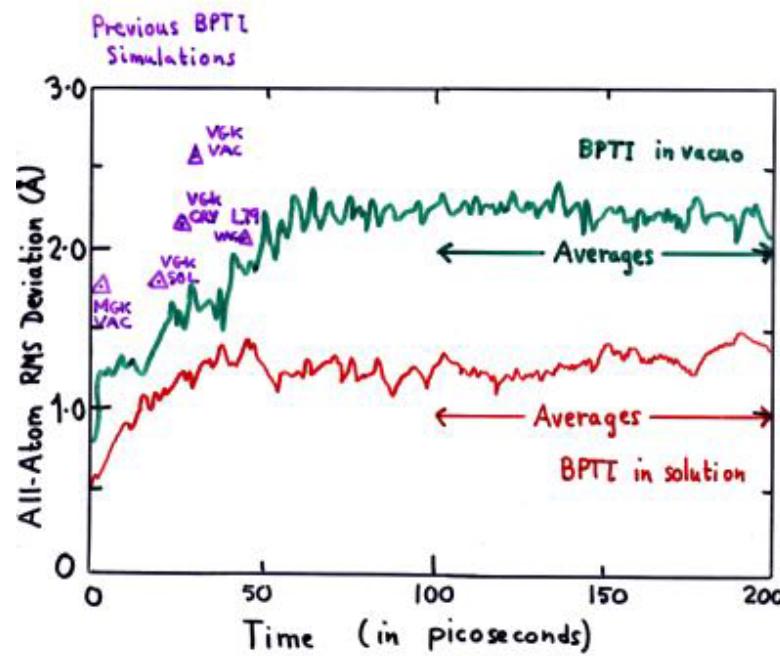
$$D = \frac{\langle \Delta r^2 \rangle}{6\Delta t}$$

# Observed RMS values

Illustration from M Levitt,  
Stanford University

## COMPARISON OF OVERALL VALUES

| Property                                      | Value    |          |          |
|---|----------|----------|----------|
|   | in vacuo | in soln. | expt.    |
| All-Atom R.M.S.<br>Deviation ( $\text{\AA}$ ) | 2.60     | 1.55     | 1.3(0.5) |
| $C^\alpha$ Fluctuation ( $\text{\AA}$ )       | 0.54     | 0.43     | 0.68     |
| Radius of Gyration ( $\text{\AA}$ )           | 10.9     | 11.5     | 11.5     |



# Other Things to Calculate

- Fraction of Native Contacts
- Percent Helix
- Radius of Gyration

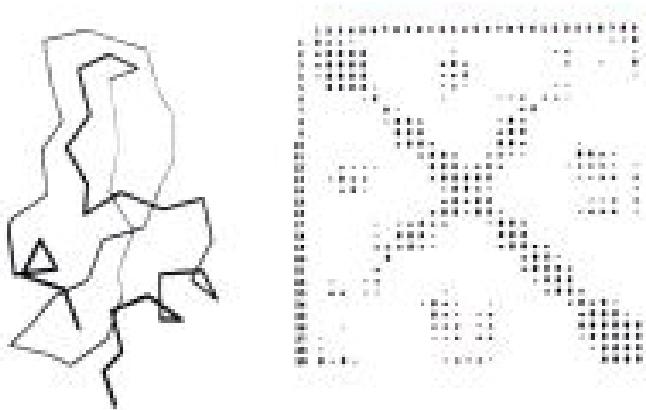
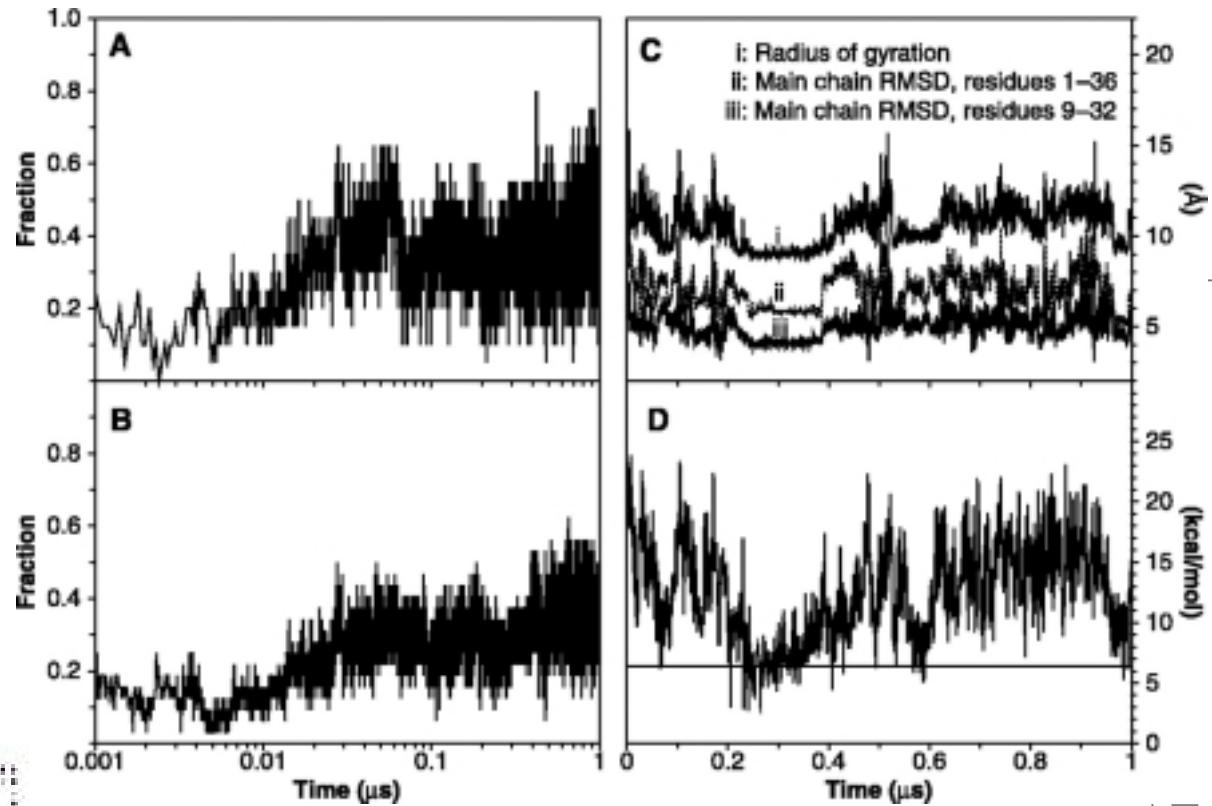


Illustration and Caption from  
Duan & Kollman (1998)



Caption: Time evolution of (A) fractional native helical content, (B) fractional native contacts, (C) R and the main chain rmsd from the native structure, and

(D) SFE of the protein. The helical content and the native contacts are plotted on a logarithmic time scale. The helical content was measured by the main chain - angle

( $60^\circ \pm 30^\circ$ ,  $40^\circ \pm 30^\circ$ ). The native contacts were measured as the number of neighboring residues present in 80% of the last 50 ns of the native simulation.

Residues are taken to be in contact if any of the atom pairs are closer than 2.8 Å, excluding residues i and i+1, which always have the contacts through main chain atoms. The SFE was calculated as described by Eisenberg and McLachlan (31) using their parameters (0.0163, 0.00637, 0.02114, 0.02376, and 0.05041, in kcal mol Å<sup>2</sup>, for the surface areas of nonpolar, polar, sulfur, charged oxygen, and charged nitrogen, respectively). The straight line represents the SFE of the native structure.

# Monitor Stability of Specific Hydrogen Bonds

Illustration from M Levitt,  
Stanford University

|                     |           | <u>HYDROGEN BONDS</u>     |                           |
|---------------------|-----------|---------------------------|---------------------------|
| Secondary Structure | O..H Pair | Stability (%)<br>in vacuo | Stability (%)<br>in soln. |
| $\beta$ -hairpin    | 35..18    | 12                        | 57                        |
|                     | 18..35    | 85                        | 63                        |
|                     | 33..20    | 71                        | 76                        |
|                     | 20..33    | 80                        | 86                        |
|                     | 31..22    | 53                        | 93                        |
|                     | 22..31    | 82                        | 87                        |
|                     | 29..24    | 72                        | 67                        |
|                     | 24..29    | 37                        | 34                        |
|                     | 45..21    | 63                        | 86                        |
|                     | 21..45    | 14                        | 42                        |
| $\alpha$ -helix     | 47..51    | 76                        | 66                        |
|                     | 48..52    | 93                        | 90                        |
|                     | 49..53    | 90                        | 98                        |
|                     | 50..54    | 78                        | 90                        |
|                     | 51..55    | 73                        | 93                        |
|                     | 52..56    | -                         | 42                        |

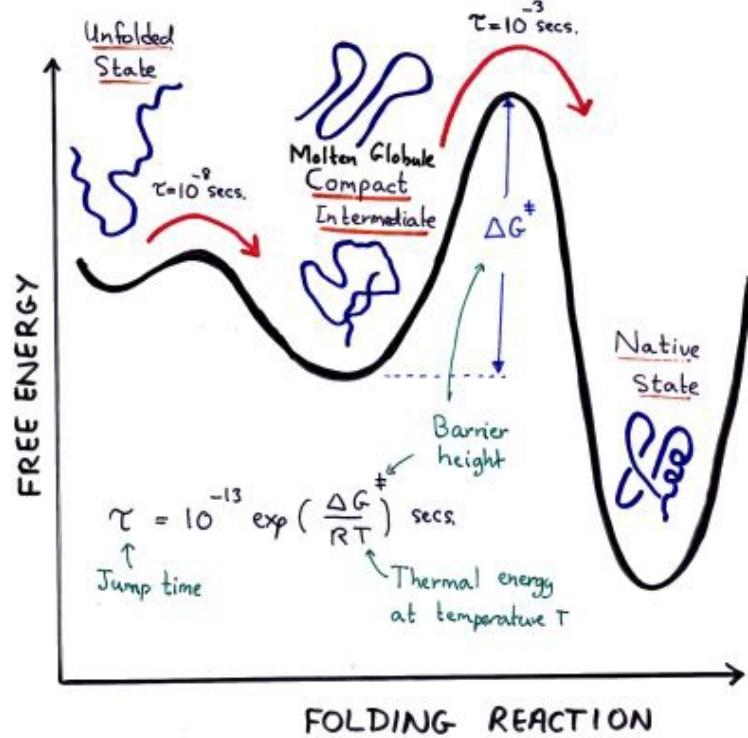
- Hydrogen bonds in solution are as strong as in vacuo
- Relative strength on position in secondary structure

# Energy Landscapes and Barriers

## Traversed in a Simulation

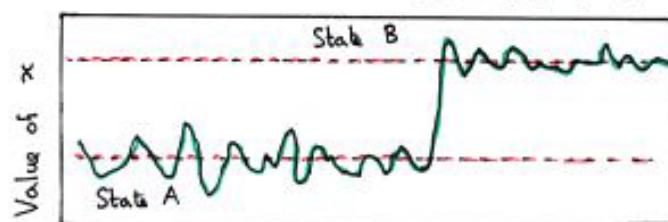
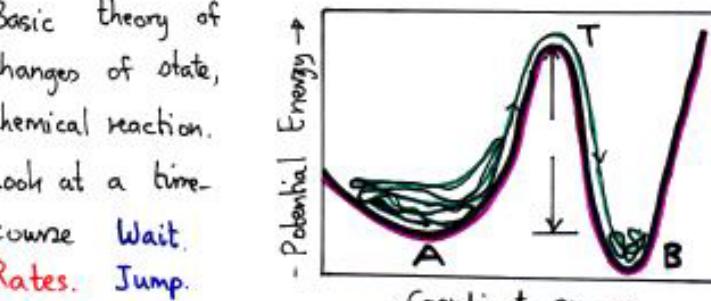
### ENERGY LANDSCAPES

Huge multi-dimensional space of changing shapes ("conformational space").



### CROSSING ENERGY BARRIERS

Basic theory of changes of state, chemical reaction.  
Look at a time-course Wait. Rates. Jump.



The actual transition from A to B is very quick (few picoseconds). What takes time is the waiting. Theory gives the average wait time in state A as

$$\tau_{A \rightarrow B} = \left( \frac{h}{kT} \right) \exp \left( + (U_T - U_A) / RT \right)$$

Planck's constant  $\left( \frac{h}{kT} \right)$   
Boltzmann constant  $\left( \frac{h}{kT} \right)$   $\approx 0.16$  picoseconds at  $T = 300^\circ K$  ( $27^\circ C$ ).  
(remember as  $h\nu_0 = kT$ ,  $\nu_0 = 208 \text{ cm}^{-1}$ )  
 $\tau_0 = 0.16 \text{ ps}$

Illustrations from M Levitt, Stanford University

---

| Motion                     | length time |           |
|----------------------------|-------------|-----------|
|                            | (Å)         | (fs)      |
| bond vibration             | 0.1         | 10        |
| water hindered rotation    | 0.5         | 1000      |
| surface sidechain rotation | 5           | $10^5$    |
| water diffusive motion     | 4           | $10^5$    |
| buried sidechain libration | 0.5         | $10^5$    |
| hinge bending of chain     | 3           | $10^6$    |
| buried sidechain rotation  | 5           | $10^{13}$ |
| allosteric transition      | 3           | $10^{13}$ |
| local denaturation         | 7           | $10^{14}$ |

## Timescales

Values from  
McCammon &  
Harvey (1987) and  
Eisenberg &  
Kauzmann

# Electrostatics Revisited: the Poisson-Boltzmann Equation

# Poisson-Boltzmann equation

- Macroscopic dielectric
  - ◊ As opposed to microscopic one as for realistic waters
- Linearized:  $\sinh \phi = \phi$ 
  - ◊ counter-ion condense
- The model
  - ◊ Protein is point charges embedded in a low dielectric.
  - ◊ Boundary at accessible surface
  - ◊ Discontinuous change to a new dielectric  
(no dipoles, no smoothly varying dielectric)

PBE Eq. Ugh!

$$\nabla \cdot [\epsilon(\vec{r}) \nabla \phi(\vec{r})] - \epsilon(\vec{r}) K(\vec{r}) \sinh[\phi(\vec{r})] - \frac{4\pi}{kT} \rho^f(\vec{r}) = 0$$

dielectric const  
IN OUT IN fixed charges  
potential ionic strength IN

# Simplifications of the Poisson-Boltzmann equation

- Laplace eq.
  - ◊  $\text{div grad } V = \rho$
  - ◊  $\text{grad } V = \mathbf{E}$  field
  - ◊ Only have divergence when have charge source

PBE Eq. Ugh !

- $\vec{\nabla} \cdot [\epsilon(\vec{r}) \vec{\nabla} \phi(\vec{r})] - \epsilon(\vec{r}) K(\vec{r}) \sinh[\phi(\vec{r})] - \frac{4\pi}{kT} \rho^f(\vec{r}) = 0$ 

The diagram shows the components of the PBE equation. It consists of four terms connected by arrows pointing towards the center of the equation. The first term is labeled 'dielectric const' with an arrow pointing to it from the left. The second term is labeled 'potential' with an arrow pointing to it from the center. The third term is labeled 'ionic strength' with an arrow pointing to it from the right. The fourth term is labeled 'fixed charges' with an arrow pointing to it from the far right. Below each term, there is a bracket indicating its sign: 'IN' under 'dielectric const', 'OUT' under 'potential', 'IN' under 'ionic strength', and 'IN' under 'fixed charges'.
- No moving ions, Constant Dielectric  $\rightarrow$  Poisson's Eq.  
$$\vec{\nabla}^2 \phi(\vec{r}) = \frac{4\pi}{kT\epsilon} \rho^f(\vec{r})$$

- Finite Difference Soln. to PDE

(PDE has deriv. WRT to 2 var.  
 (ODE like Newton's eq. has deriv. WRT to 1 var.)

$$\nabla^2 \varphi(r) = \frac{4\pi}{kT\varepsilon} \rho(r)$$

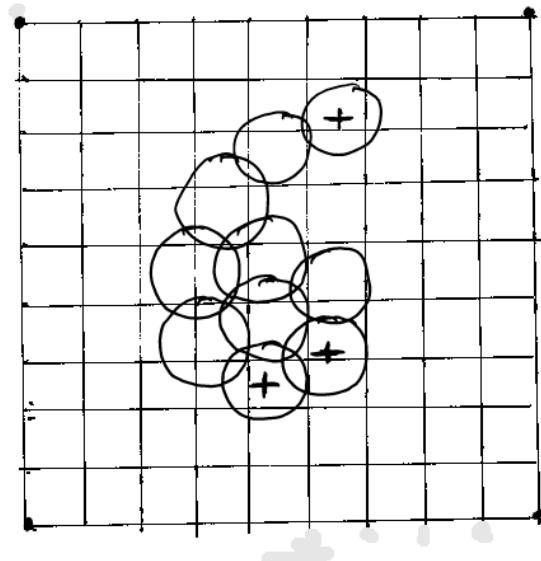
const.

$$\text{in 2D } \frac{\partial \varphi}{\partial x^2} + \frac{\partial \varphi}{\partial y^2} = C \rho(r)$$

$$\frac{\partial^2 \varphi}{\partial x^2} = \frac{\partial}{\partial x} \left( \frac{\partial \varphi}{\partial x} \right) = \frac{\partial}{\partial x} \left( \frac{V_{j+1} - V_j}{\Delta} \right) = \frac{(V_{j+1} - V_j) - (V_j - V_{j-1})}{\Delta^2}$$

$(\Delta x = \Delta, \Delta \varphi = V_{j+1} - V_j)$

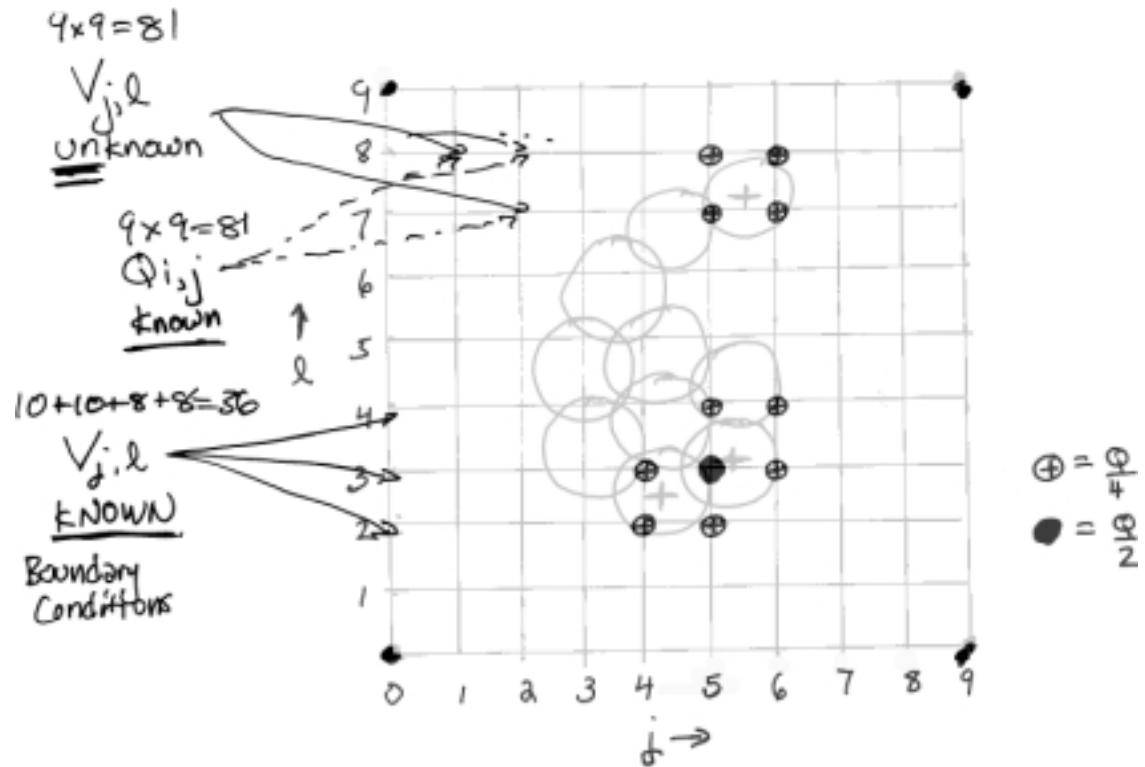
$$V_{j+1,l} + V_{j-1,l} + V_{j,l+1} + V_{j,l-1} - 4V_{j,l} = \Delta^2 C Q_{j,l}$$



bb.yale.edu

# Protein on a Grid

For intuition ONLY -- Don't  
need to know in detail!!



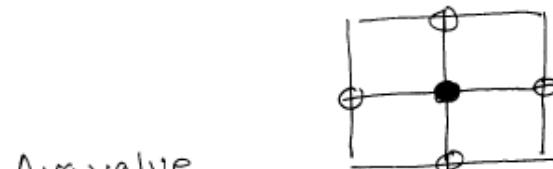
# Demand Consistency on the Grid

- $V_{j+1,l} + V_{j-1,l} + V_{j,l+1} + V_{j,l-1} - 4V_{j,l} = \Delta^2 C Q_{j,l}$
- System of Equations  $\rightarrow$  solve for unknown  $V_{j,l}$
- Matrix Inversion in Finite Diff. method

Relaxation: Deviation from consistency should vanish at  $t \rightarrow \infty$

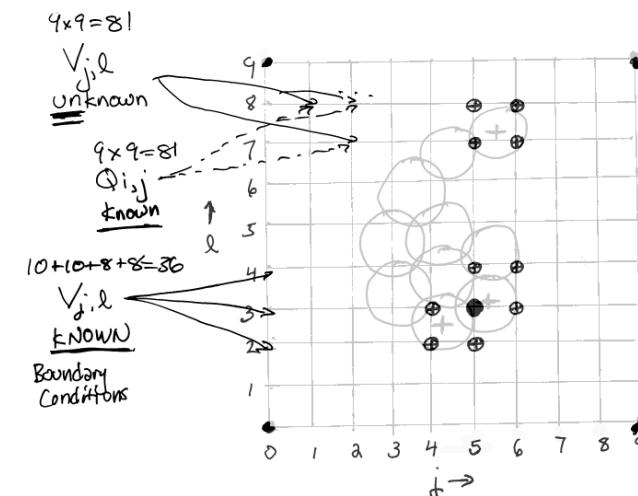
$$\nabla^2 V - 4\pi \rho = \left( \frac{\partial V}{\partial t} \right) \underset{t \rightarrow \infty}{\sim} 0$$

$$V_{j,l}^{t+1} \leftarrow V_{j,l}^t + \Delta t \left( \frac{V_{j+1,l}^t + V_{j-1,l}^t + V_{j,l+1}^t + V_{j,l-1}^t - 4V_{j,l}^t}{\Delta^2} - Q_{j,l} \right)$$



Avg value  
at center (●) is  
avg. value at 4 outside nodes (⊕)  
plus charge at center

For intuition ONLY  
-- Don't need to  
know in detail!!

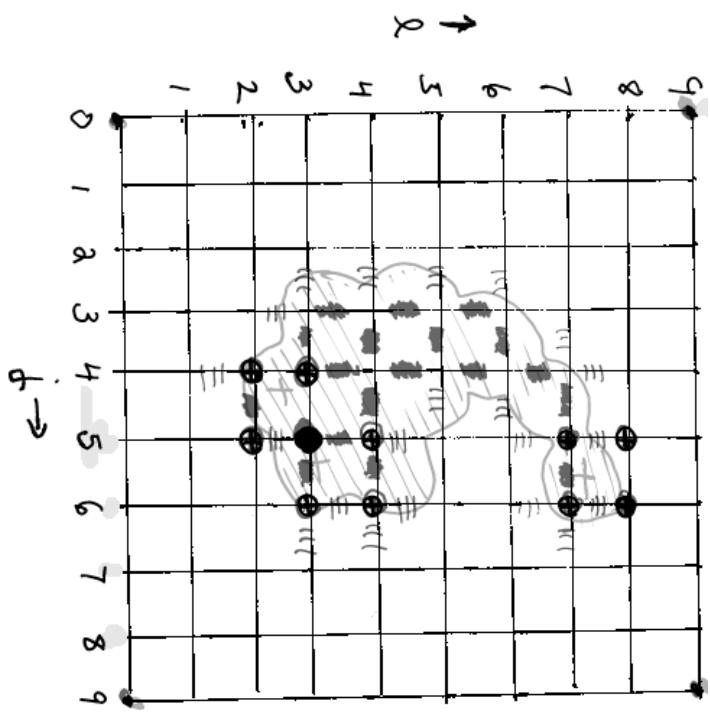


# Adding a Dielectric Boundary into the Model

$$\nabla \cdot (\epsilon(r) \nabla \varphi) \Rightarrow$$

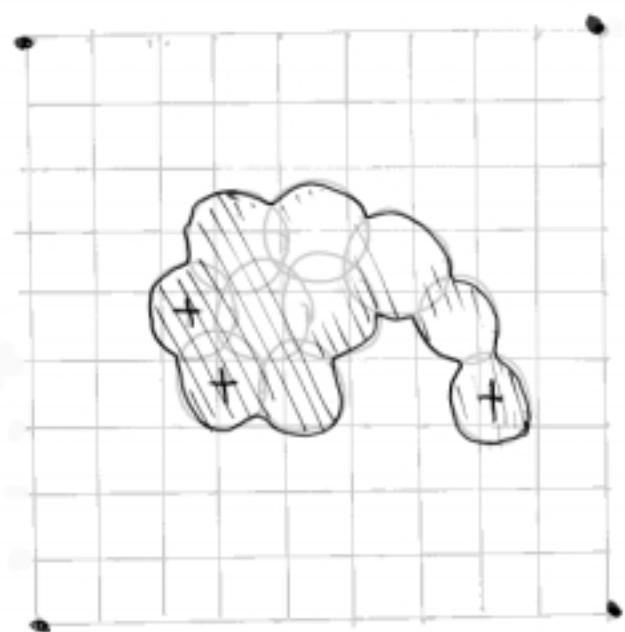
$$\frac{1}{r} (\epsilon_{j \rightarrow j+1} (\nabla_{j+1} - \nabla_j) -$$

$$\epsilon_{j \rightarrow j} (\nabla_j - \nabla_{j-1}))$$



$$\oplus = \frac{\Theta + \bar{\Theta}}{2}$$

$$- = \frac{\Theta - \bar{\Theta}}{2}$$



$\epsilon = 2$  inside  
 $\epsilon = 80$  outside

# Electrostatic Potential of Thrombin

The proteolytic enzyme Thrombin (dark backbone worm) complexed with an inhibitor, hirudin (light backbone worm). The negatively charged (Light gray) and positively charged (dark gray) sidechains of thrombin are shown in bond representation.

Graphical analysis of electrostatic potential distributions often reveals features about the structure that complement analysis of the atomic coordinates. For example, LEFT shows the distribution of charged residues in the binding site of the proteolytic enzyme thrombin. RIGHT shows the resulting electrostatic potential distribution on the protein surface. The basic (positive) region in the fibrinogen binding, while it could be inferred from close inspection of the distribution of charged residues in TOP, is more apparent in the potential distribution.

Solvent accessible surface of thrombin coded by electrostatic potential (dark: positive, light: negative). Hirudin is shown as a light backbone worm. Potential is calculated at zero ionic strength.

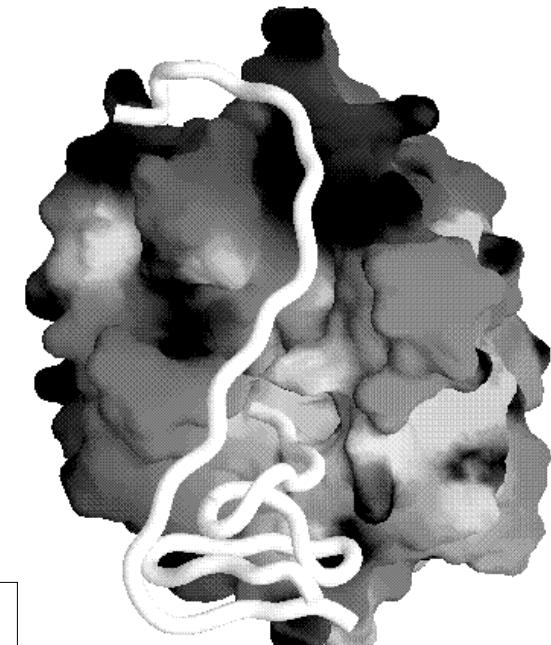


Illustration Credit: Sharp (1999)  
Text captions also from Sharp (1999)

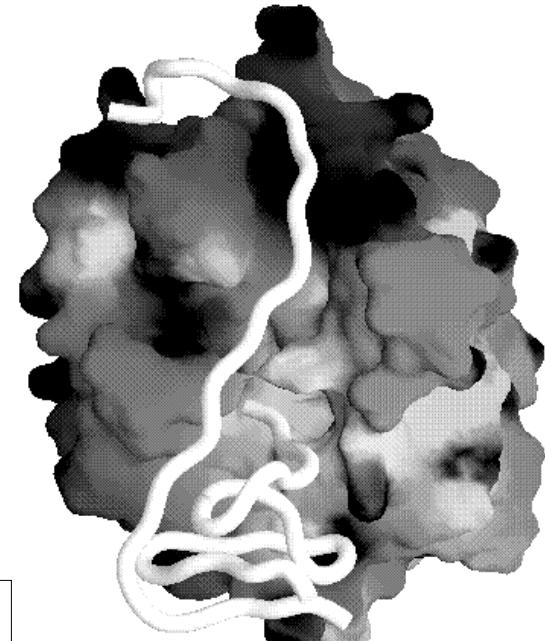
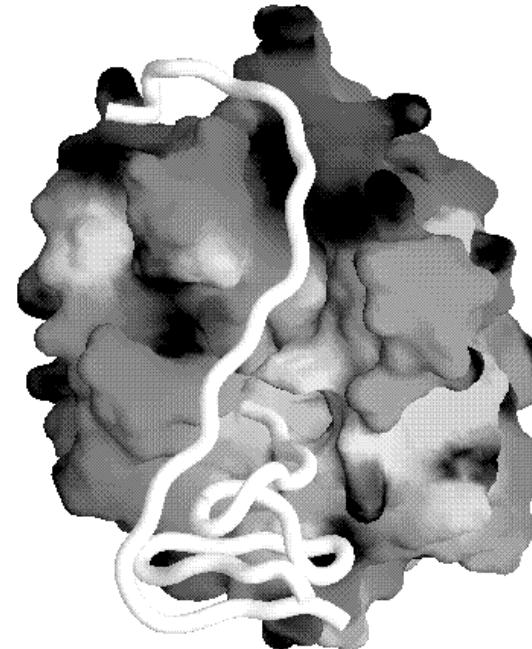
# Increasing Ionic Strength

Solvent accessible surface of thrombin coded by electrostatic potential (dark: positive, light: negative). Hirudin is shown as a light backbone worm. Potential is calculated at physiological ionic strength (0.145M)

TOP shows the effect of increasing ionic strength on the potential distribution, shrinking the regions of strong potential in comparison to BOTTOM.

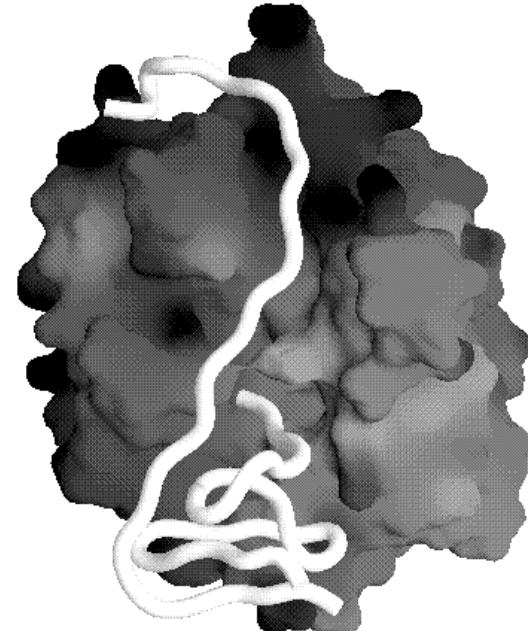
Solvent accessible surface of thrombin coded by electrostatic potential (dark: positive, light: negative). Hirudin is shown as a light backbone worm. Potential is calculated at zero ionic strength.

Illustration Credit: Sharp (1999)  
Text captions also from Sharp (1999)



# Increasing Dielectric

Solvent accessible surface of thrombin coded by electrostatic potential (dark: positive, light: negative). Hirudin is shown as a light backbone worm. Potential is calculated using the same polarizability for protein and solvent.



TOP is calculated assuming the same dielectric for the solvent and protein. The more uniform potential distribution compared to BOTTOM shows the focusing effect that the low dielectric interior has on the field emanating from charges in active sites and other cleft regions.

Solvent accessible surface of thrombin coded by electrostatic potential (dark: positive, light: negative). Hirudin is shown as a light backbone worm. Potential is calculated at zero ionic strength.

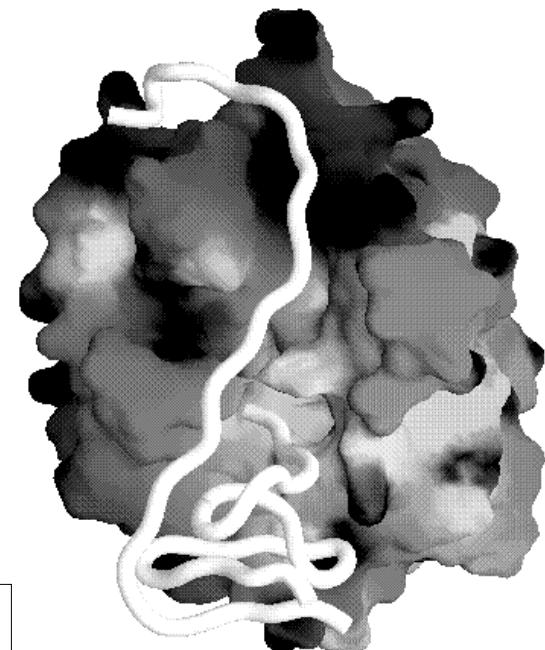


Illustration Credit: Sharp (1999)  
Text captions also from Sharp (1999)

# pKa shifts

Charge transfer processes are important in protein catalysis, binding, conformational changes and many other functions. The primary examples are acid-base equilibria, electron transfer and ion binding, in which the transferred species is a proton, an electron or a salt ion respectively. The theory of the dependence of these three equilibria within the classical electrostatic framework can be treated in an identical manner, and will be illustrated with acid-base equilibria. A titratable group will have an intrinsic ionization equilibrium, expressed in terms of a known intrinsic  $pK^0a$ . Where  $pK^0a = -\log_{10}(K^0a)$ ,  $K^0a$  is the dissociation constant for the reaction  $H^+A = H^++A$  and A can be an acid or a base. The  $pK^0a$  is determined by all the quantum chemical, electrostatic and environmental effects operating on that group in some reference state. For example a reference state for the aspartic acid side-chain ionization might be the isolated amino acid in water, for which  $pK^0a = 3.85$ . In the environment of the protein the pKa will be altered by three electrostatic effects. The first occurs because the group is positioned in a protein environment with a different polarizability, the second is due to interaction with permanent dipoles in the protein, the third is due to charged, perhaps titratable, groups. The effective pKa is given by (where the factor of  $1/2.303kT$  converts units of energy to units of pKa):

$$pKa = pK^0a + (\Delta\Delta G_{rf} + \Delta\Delta G_{perm} + \Delta\Delta G_{tit})/2.303kT$$

Text block from  
Sharp (1999)

1. Desolvation,  
Rx Field

2. Permanent  
Dipoles

3. Other  
Charges

# pKa

## continued I

Text block from  
Sharp (1999)

### 1. Desolvation, Rx Field

The first contribution,  $\Delta\Delta G^{\text{rf}}$ , arises because the completely solvated group induces a strong favorable reaction field (See section 22.3.2.3) in the high dielectric water, which stabilizes the charged form of the group (The neutral form is also stabilized by the solvent reaction field induced by any dipolar groups, but to a lesser extent). Desolvating the group to any degree by moving it into a less polarizable environment will preferentially destabilize the charged form of that group, shifting the pKa by an amount

$$\Delta\Delta G^{\text{rf}} = \frac{1}{2} \sum_i (q_i^d \Delta\phi_i^{\text{rf},d} - q_i^p \Delta\phi_i^{\text{rf},p}) \quad (12)$$

where  $q_i^p$  and  $q_i^d$  are the charge distributions on the group.  $\Delta\phi_i^{\text{rf},p}$  and  $\Delta\phi_i^{\text{rf},d}$  are the changes in the group's reaction potential upon moving it from its reference state into the protein, in the protonated (superscript p) and deprotonated (superscript d) forms respectively, and the sum is over the group's charges.

The contribution of the permanent dipoles is given by

$$\Delta\Delta G^{\text{tit}} = \sum_i (q_i^d - q_i^p) \phi_i^{\text{perm}} \quad (13)$$

where  $\phi_i^{\text{perm}}$  is the interaction potential at the  $i$ 'th charge due to all the permanent dipoles in the protein, including the effect of screening. It is observed that intrinsic pKa's of groups in proteins are rarely shifted by more than 1 pKa unit indicating that the effects of desolvation are often compensated to a large degree by the  $\Delta\Delta G^{\text{perm}}$  term.

### 2. Permanent Dipoles

# pKa continued II

The final term accounts for the contribution of all the other charge groups:

$$\Delta\Delta G^{\text{titr}} = \sum_i (q_i^d <\phi_i>_{\text{pH}, c, \Delta V}^d - q_i^p <\phi_i>_{\text{pH}, c, \Delta V}^p) \quad (14)$$

## 3. Other Charges

Text block from Sharp (1999)

where  $\langle \phi_i \rangle$  is the mean potential at group charge  $i$  from all the other titratable groups. The charge state of the other groups in the protein depend in turn on their intrinsic "pKa's", on the external pH if they are acid-base groups, the external redox potential  $\Delta V$  if they are redox groups, and the concentration of ions,  $c$ , if they are ion binding sites, as indicated by the subscript on  $\langle \phi_i \rangle$ . Moreover, the charge state of the group itself will affect the equilibrium at the other sites. Because of this linkage, exact determination of the complete charged state of a protein is a complex procedure. If there are  $N$  such groups, the rigorous approach is to compute the titration state partition function by evaluating the relative electrostatic free energies of all  $2^N$  ionization states for a given set of pH,  $c$ ,  $\Delta V$ . From this one may calculate the mean ionization state of any group as a function of pH,  $\Delta V$  etc. For large  $N$  this becomes impractical, but various approximate schemes work well, including a Monte-Carlo procedure

# Water Simulation and Hydrophobicity

# Simulating Liquid Water

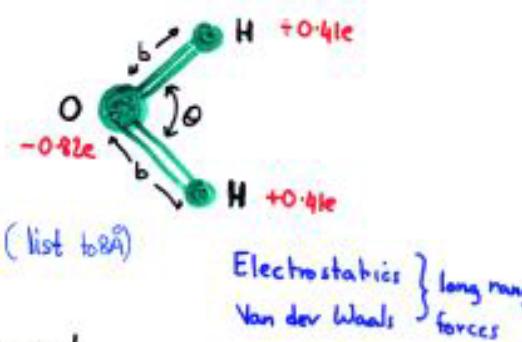
## SIMULATING LIQUID WATER

- Very simple model

- 3 interaction centers

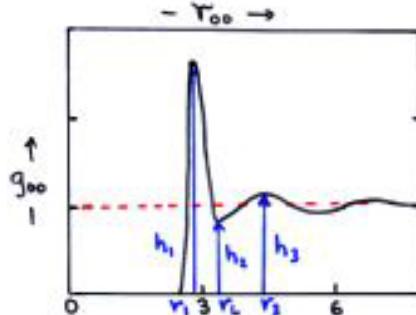
- Completely flexible

- Smooth cutoff at  $6\text{ \AA}$  (list to  $8\text{ \AA}$ )



- Good fit to experiment

| Property   | Experiment   | Simulation                              |
|--|--|---|
| Potential energy (kcal/mol)                        | -9.2   | -9.5                                    |
| Pressure (atmospheres)                             | [1]  | [61]                                    |
| Classical Specific Heat (cal/ $^{\circ}\text{K}$ ) | 27   | 26                                      |
| Diffusion Constant ( $\text{\AA}^2/\text{ps}$ )    | 0.23   | 0.22                                    |
| Rotational Relaxation Time (ps)                    | 2  | 1.6                                     |
| Radial Distribution Function                       | $r_1$<br>$h_1$<br>$r_2$<br>$h_2$<br>$r_3$<br>$h_3$ | 2.8<br>2.5<br>3.3<br>0.8<br>4.6<br>1.11 |

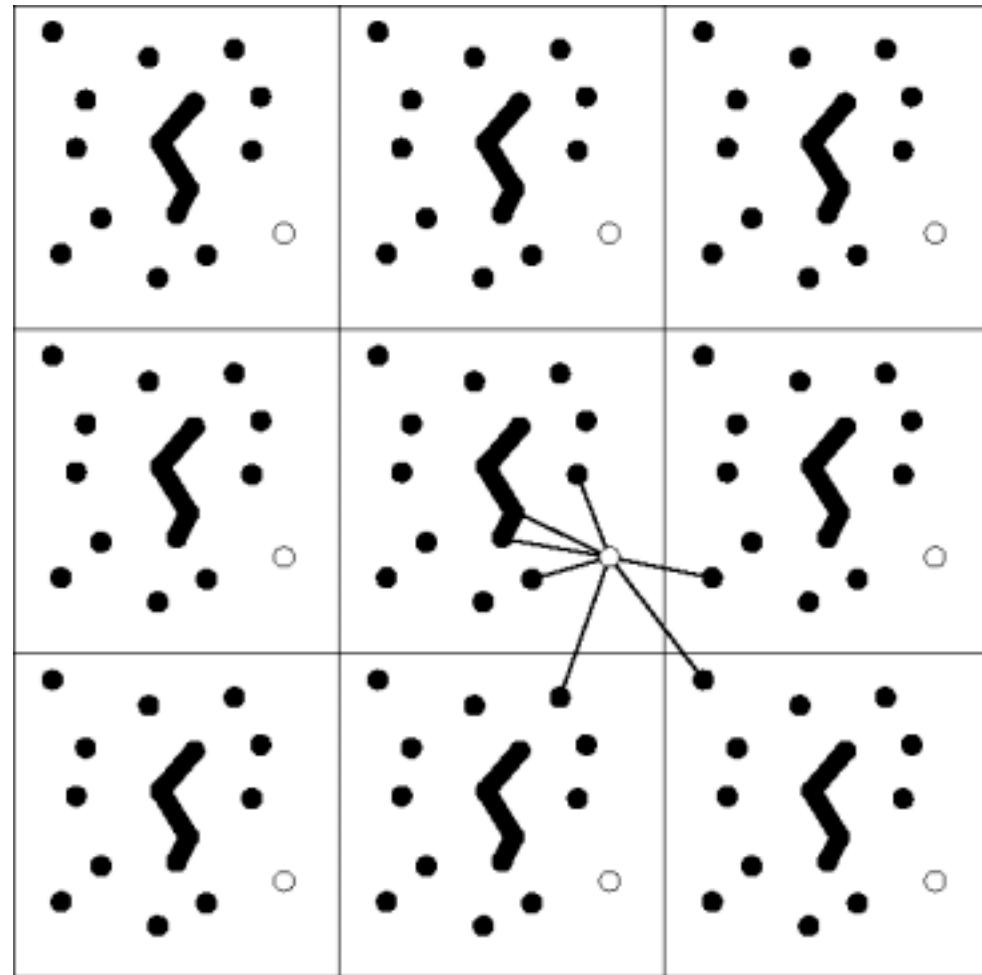


\* Calibration error fixed after  
15 years of experiment

Illustrations from  
M Levitt, Stanford  
University

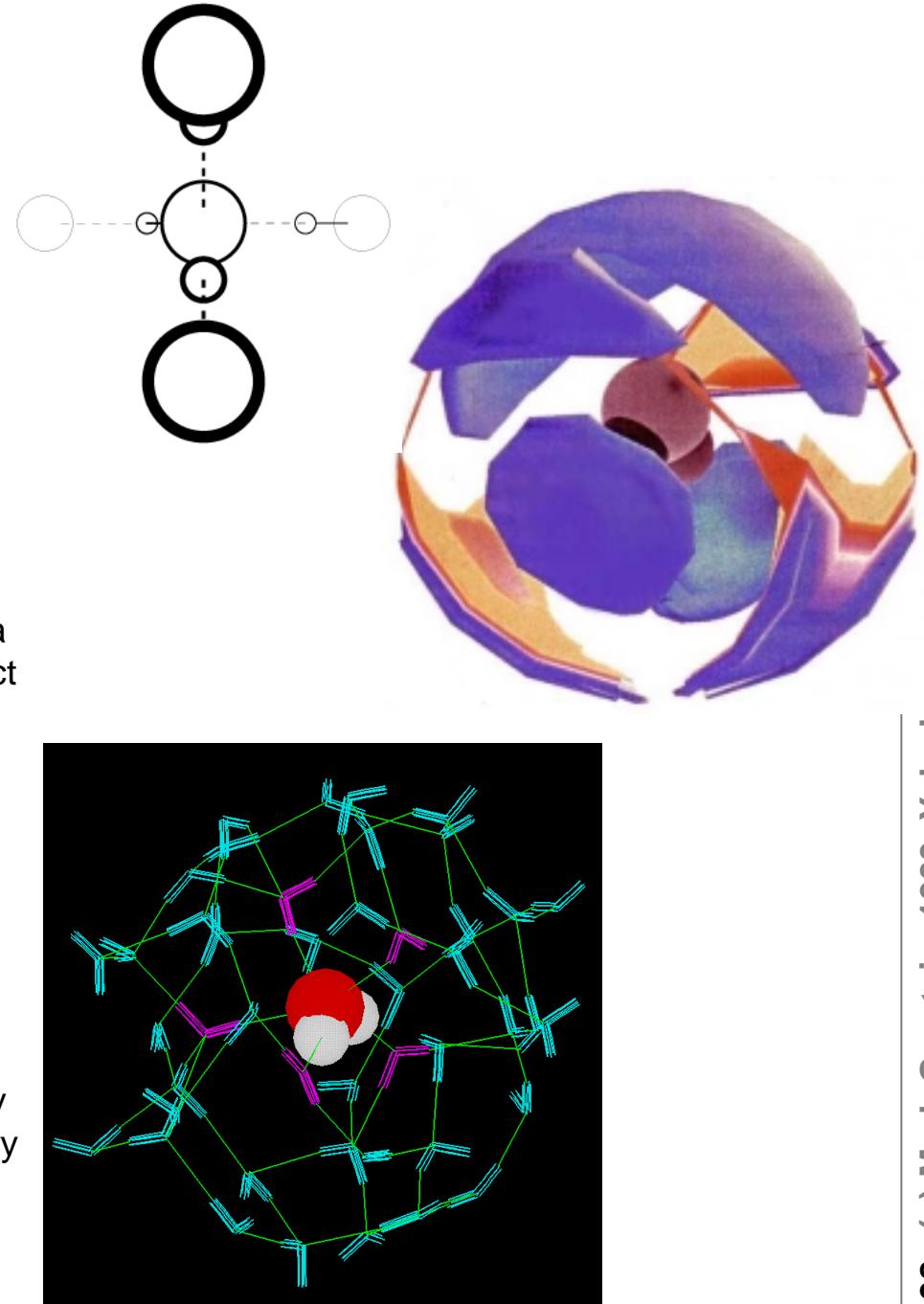
# Periodic Boundary Conditions

- Make simulation system seem larger than it is
- Ewald Summation for electrostatics (Fourier transform)



# Tetrahedral Geometry of Water

HYDROGEN BONDS give water its unique properties. The hydrogen bond is a consequence of the electrical attraction between the positively charged hydrogen on one water molecule ( $H_1$ ) and the negatively charged oxygen on another water molecule ( $O'$ ). The electrostatic repulsion between this oxygen and the oxygen that the hydrogen is covalently bonded to ( $O$ ) gives the hydrogen bond a nearly linear geometry. Each water molecule can act as a donor of two hydrogen bonds to neighboring water oxygens. Each water can also accept two hydrogen bonds. This double-donor, double-acceptor situation naturally tends to favor a tetrahedral geometry with four waters around each water oxygen, as shown. Ice has this perfect tetrahedral geometry. However, in water, the tetrahedral geometry is distorted, and it is possible for a water molecule to accept or donate more than two hydrogen bonds (which are consequently highly distorted). Such distortions of tetrahedral geometry are shown, which is taken from a frame in a simulation. Note that the central water molecule accepts three hydrogen bonds.



# Hydrophobicity Arises Naturally in Simulation

- Add no hydrophobic Effect
  - ◊ This arises naturally from entropic effects during the simulation

Mixing is a spontaneous process: a substance will naturally dissolve in water unless there are manifestly unfavorable interactions between it and water. Scientists usually discuss the favorableness of particular interactions in terms of the energy associated with the intermolecular forces. Almost always there are at least some energetically favorable dispersion interactions between the solute and the water. However, the more salient issue is how the interaction between a solute and a water molecule *compares* in strength to the interaction between two water molecules or between two solute molecules. For instance, a polar molecule such as glucose is able to make comparable hydrogen bonds to water as water molecules can make with each other. Thus, there are no unfavorable interactions preventing it from dissolving and it is very soluble.

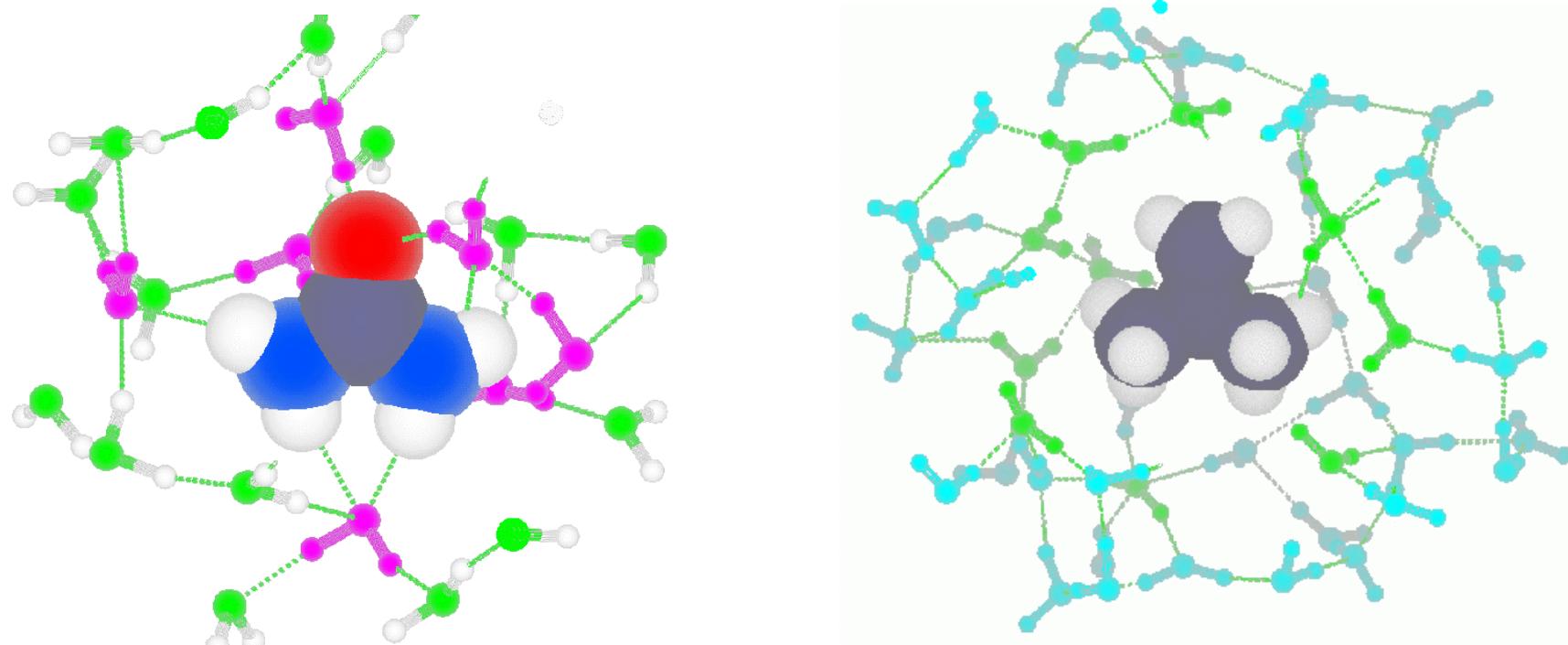
In contrast, water molecules are not able to hydrogen bond to methane, an insoluble, non-polar solute. They would rather interact with each other. The methane molecules, moreover, can favorably interact with each other through attractive dispersion forces. One can see how this situation leads to methane molecules trying to minimize their *relatively* unfavorable interactions with water molecules. An obvious way they can do this is by clumping together, aggregating, and coming out solution. Such aggregation of non-polar solutes in water is often called the *hydrophobic effect* and, as we shall, it is very important in macromolecular structure.

In terms of water structure at room temperature, the relatively unfavorable interaction between water and methane induces each water molecule next to methane to “turn away” from it and hydrogen bond to neighboring water molecules. If one of these turned water molecules manages to keep itself correctly oriented over time, it will have will not have to sacrifice any of its usual four to five hydrogen bonds. This brings up an interesting paradox: From the standpoint of favorable interactions, or energy in more formal terminology, water has not paid any price in solvating the methane. Consequently, there appears to be no energetic reason for methane to be insoluble in water.

This paradox is resolved by entropy. According to one way of thinking, entropy reflects the number of possible states a molecule can exist in. Thus, the more states a water molecule can exist in, the better its situation is entropically, and if a solute “pins down” a water molecule or restricts its freedom of motion, it is entropically unfavorable. All solutes restrict the freedom of motion of water molecules to some degree, but this is particularly true for a non-polar solute, such as methane. Thus, since turning away from methane “pins down” each water molecule slightly, the price of hydrating this non-polar solute is paid indirectly in terms of entropy and not directly in terms of energy.

The hydrophobic effect is currently receiving intense scrutiny from simulation and experiment. The picture that is emerging is somewhat more complicated than the simplified account presented here since at high temperatures, hydrophobic hydration is still unfavorable but for energetic and not entropic reasons. Nevertheless, irrespective of whether the price is paid in terms of energy or entropy, the hydrophobic effect is fundamentally caused by the *relatively* unfavorable interactions between water and hydrophobic solutes.

# Different Behavior of Water around Hydrophobic and Hydrophilic Solutes



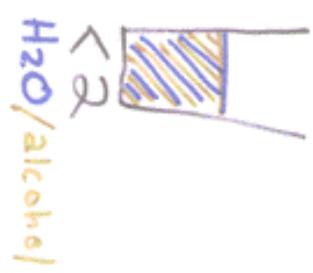
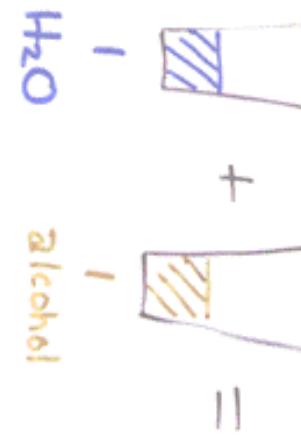
POLAR AND NON-POLAR SOLUTES have very different effects on water structure. We show two solutes that have the same Y-shaped geometry but different partial charges. The polar solute, urea (*left*), has partial charges on its atoms. Consequently, it is able hydrogen-bond to water molecules and to fit right into the water hydrogen-bond network. In contrast, the non-polar solute, isobutene (*right*), does not have (substantial) partial charges on any of its atoms. It, thus, can not hydrogen-bond to water. Rather, the water molecules around it “turn away” and interact strongly only with other water molecules, forming a sort of hydrogen-bond “cage” around the isobutene.

# Consequences of Hydrophobic Hydration and “Clathrate” Formation

- Hydrophobic hydration is unfavorable (G) but the reason is different at different T
  - ◊ entropically (S) unfavorable at low temperatures because of ordering
  - ◊ enthalpically (H) unfavorable at high temperatures because of unsatisfied H-bonds
- Volume of mixing is negative
- Compressibility
- High heat capacity of hydrophobic solvation
  - ◊ Signature of hydrophobic hydration
  - ◊ Hydration creates new temperature “labile” structures

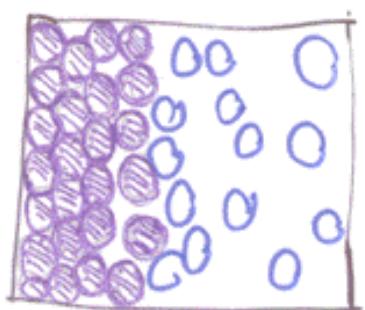
# Ways of Rationalizing Packing

(-)



(+)

TIGHT CORE  
( $>$  organic crystals)



• 4  
• 6  
• 8

## Compare standard volumes with amino acids CRYSTAL volumes

Example residue volume: Leu ( $\text{\AA}^3$ )

|     |                               |      |
|-----|-------------------------------|------|
| 1 - | Residue in the protein core   | 165  |
| 2 - | VDW envelope                  | 128  |
| 3 - | Absolute packing efficiency   | 78 % |
| 3 - | Sidechain in the protein core | 101  |
| 4 - | Sidechain in a.a. crystal     | 110  |
| 4 - | Sidechain in solution         | 107  |

Example atomic volume:  $-\text{CH}_2-$  ( $\text{\AA}^3$ )

|                    |      |
|--------------------|------|
| Protein core       | 23.5 |
| In solution        | 26.5 |
| In organic solvent | 29.0 |

Overall  
comparison to  
crystal volumes

- 3- • 4% less on avg.  
• Exceptionally  
tight core  
packing



## Compare Standard Core Volumes

### with Amino Acid Solution Volumes

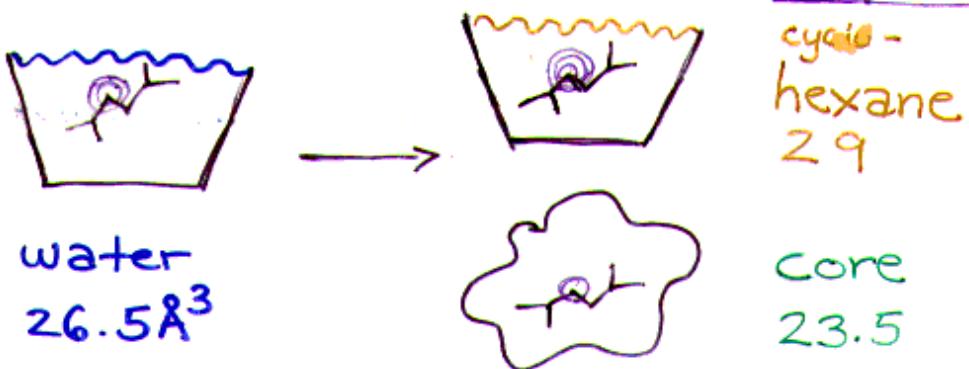
$$\sqrt{V_{\text{SIDECHAIN}}^{\text{SOLUTION}}} - \sqrt{V_{\text{CORE}}^{\text{CORE}}} \quad (\text{Cohn et al. '34}) \\ (\text{Rao et al. '84})$$

= + 4 ALIPHATICS (A Y L I P)

= ~ 0 POLARS, AROMATICS (M C F Y  
W S T)

= - 7 CHARGED, AMIDE (H N D Q  
E R K)

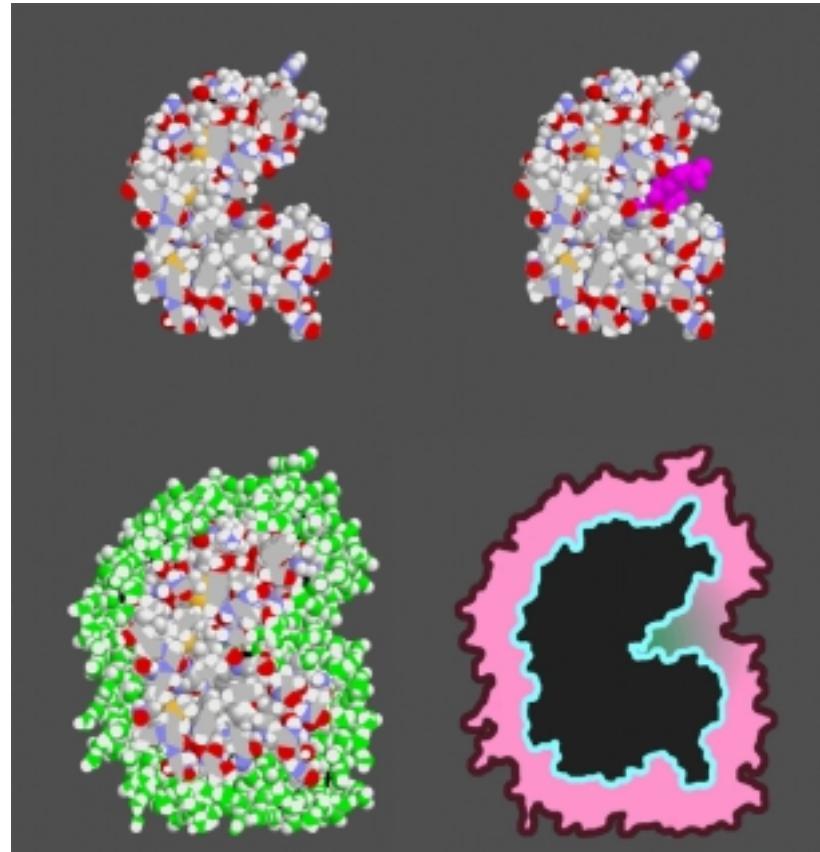
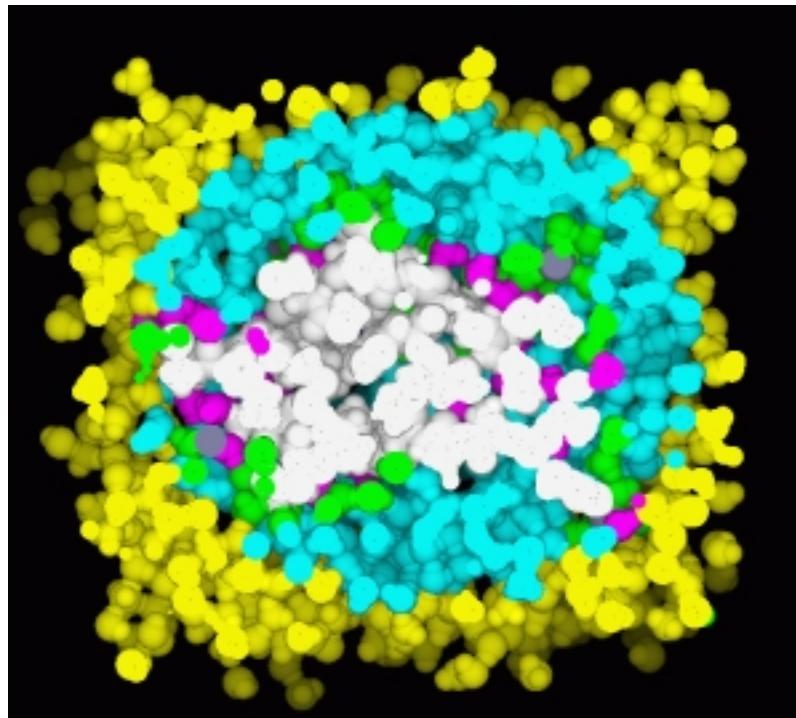
SOLUTION-TRANSFER Models Predict  
Opposite Result for Aliphatics



# Water around Hydrophobic Groups on protein surface is more Compressible

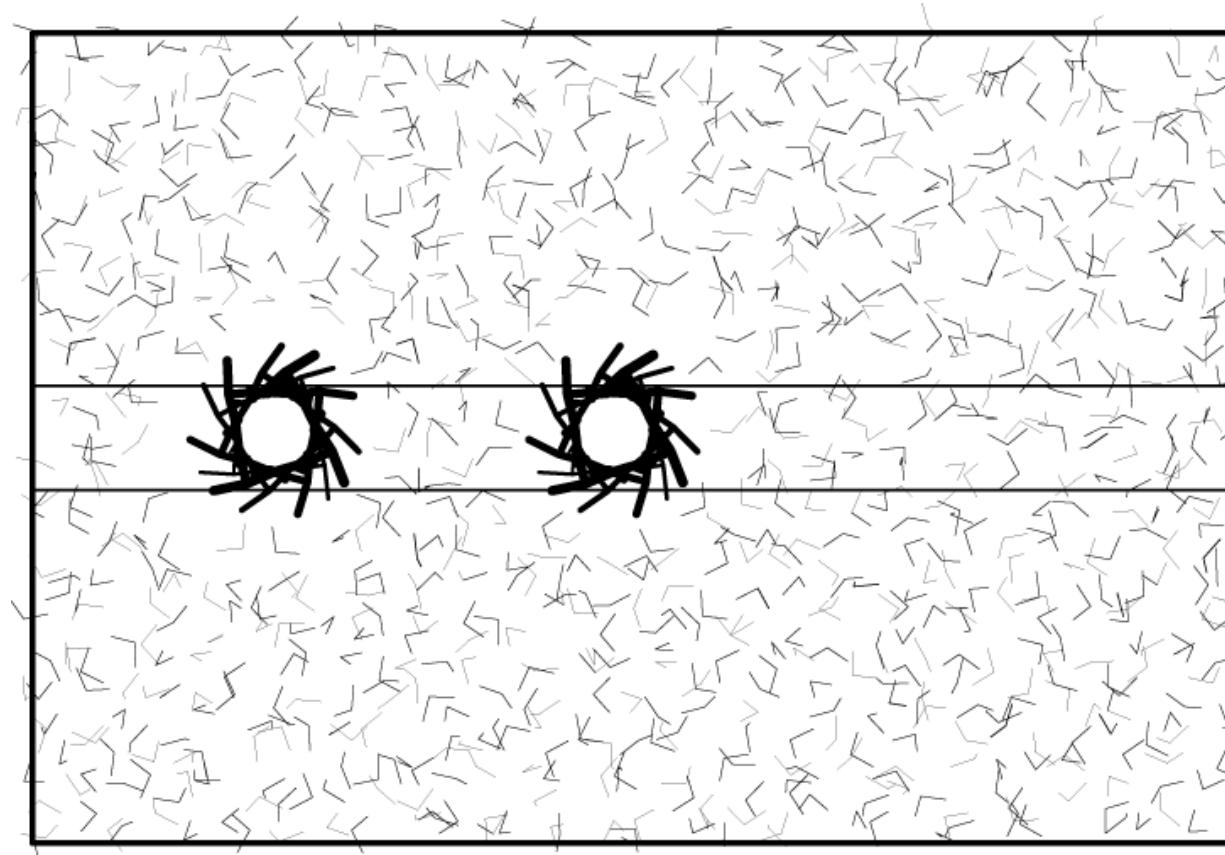
- Fluctuations in polyhedra volume over simulation related to compressibility
  - ◊ Same way amplitude of a spring is related to spring constant
  - ◊ Rigorous for NPT only, approximately true for part of NVE
- Simulation Results (avg. fluctuations as %SD and compressibility)
  - ◊ Protein core                                    9.7 %                            .14
  - ◊ Protein surface                                11.7 %                            .29
  - ◊ Water near protein                            13.2 %                            .50
  - ◊ Bulk water                                      11.9 %                            .41
  - ◊ Consistent with more variable packing at protein surface
- Results verified by doing high-pressure simulation (5000 atm, 10000 atm)
  - ◊ Allows calculation of compressibility from definition

# Interaction Between Water and the Protein Surface



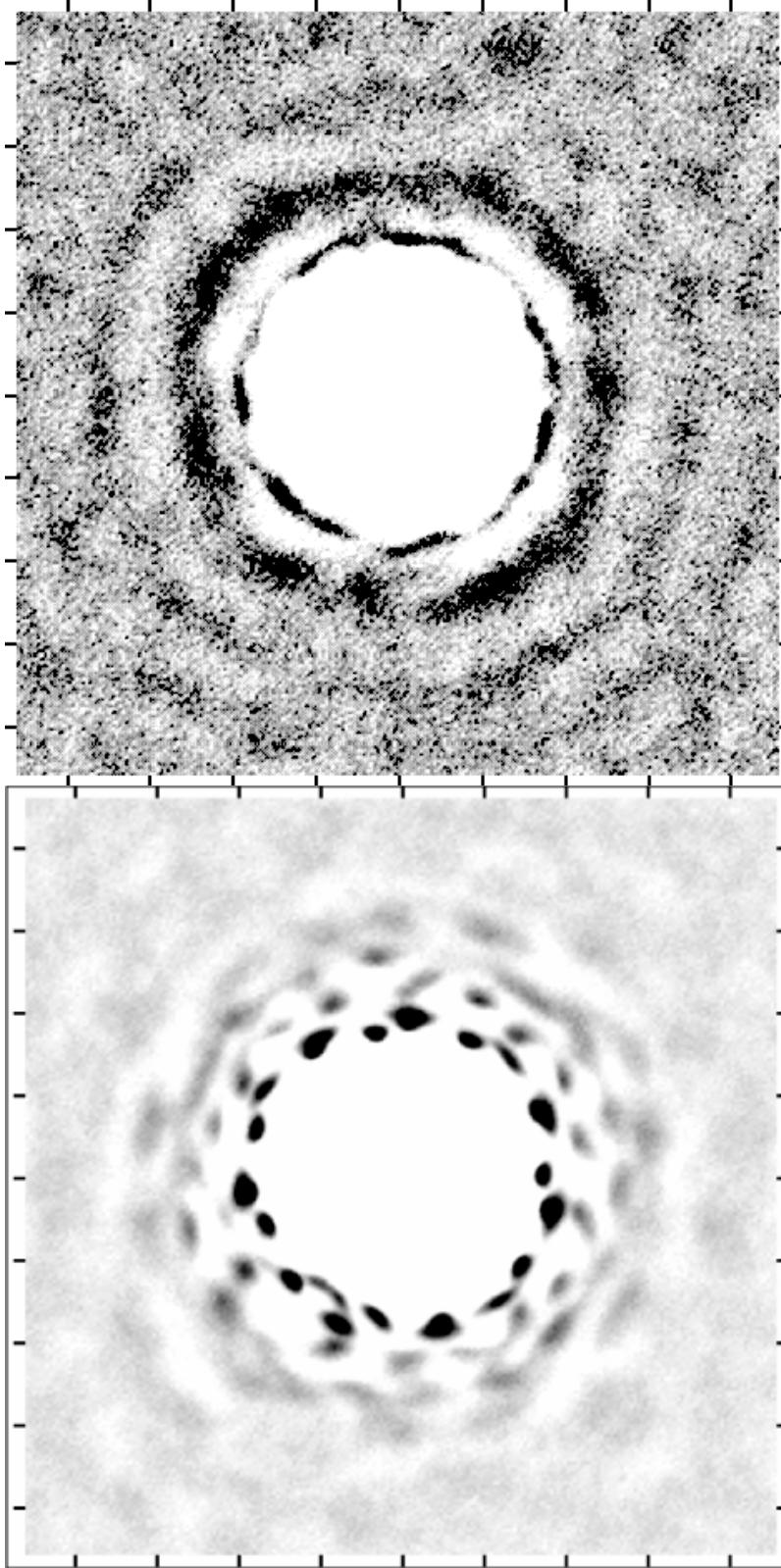
THE PROTEIN SURFACE presents a very interesting interface from the point of view of water structure since it has a very irregular shape and has polar and non-polar atoms juxtaposed in close proximity. A slice through one frame of a simulation of water around a protein is shown. The protein is shown with white atoms in the center. Water molecules strongly interacting with polar and non-polar atoms on the protein surface are shown in magenta and green, respectively. Water molecules weakly interacting with protein are shown in blue. The “region of influence” of the protein extends to roughly the second layer of water molecules. After that the water molecules are not strongly perturbed by the protein. These unperturbed, “bulk” water molecules are shown in yellow. Also, at the center of the protein one can see two buried waters (*magenta*).

# Simple Two Helix System



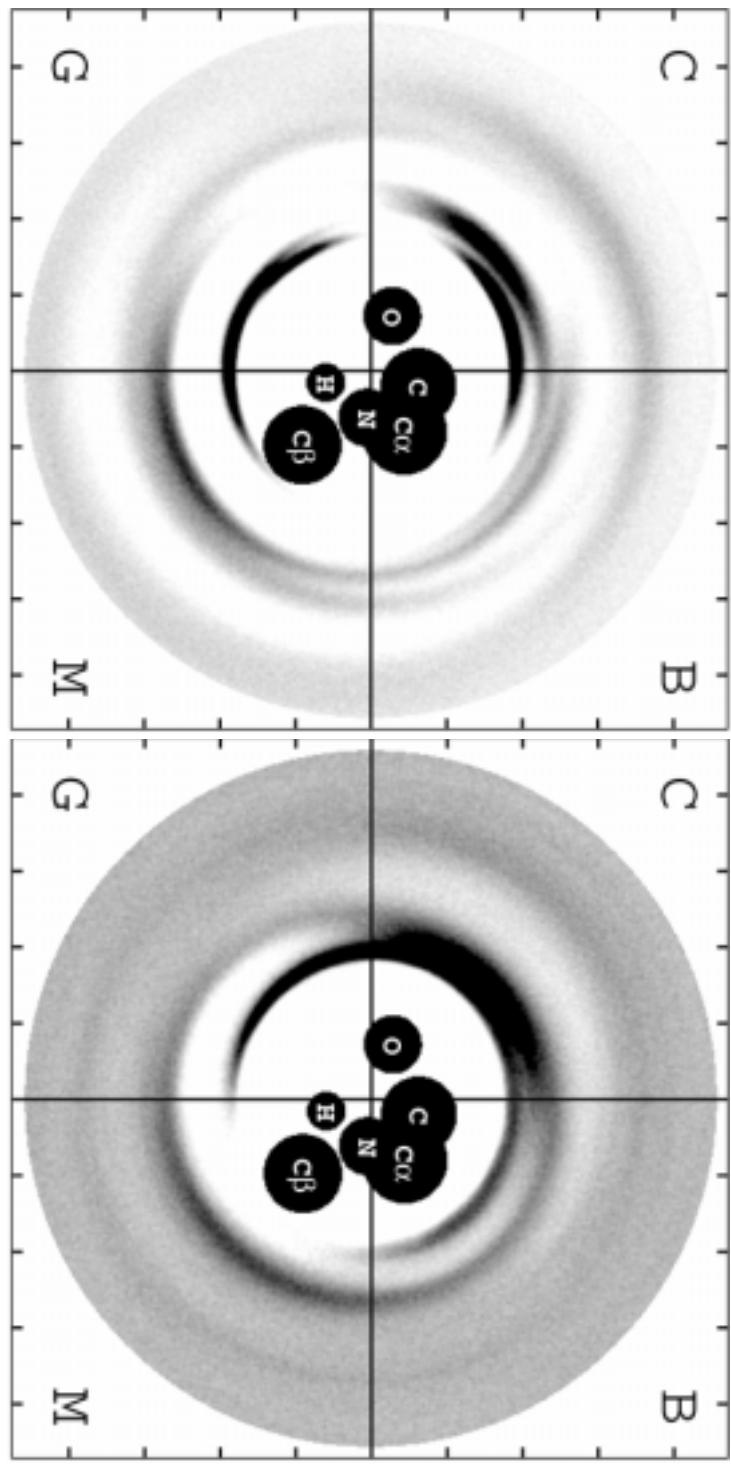
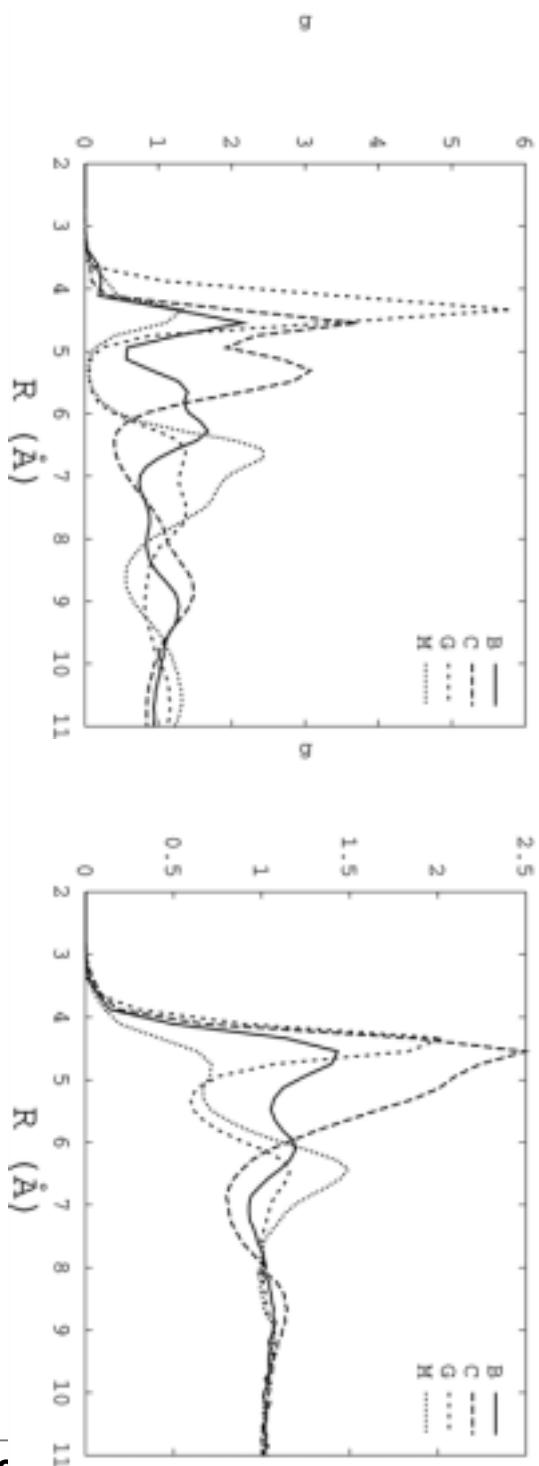
- Number density
  - ◊  $g =$  Normal  
water,straight &  
helical projections
  - ◊ For usual RDF  
“volume elements”  
are concentric  
spherical shells
  - ◊ Here, they are tiny  
vertical columns and  
helices  
perpendicular to  
page
  - ◊ More intuition about  
groove expansion
- Compare water  
packing with that of  
simple liquid (“re-  
scaled Ar”)

Second Solvent Shell:  
Water v LJ Liquid



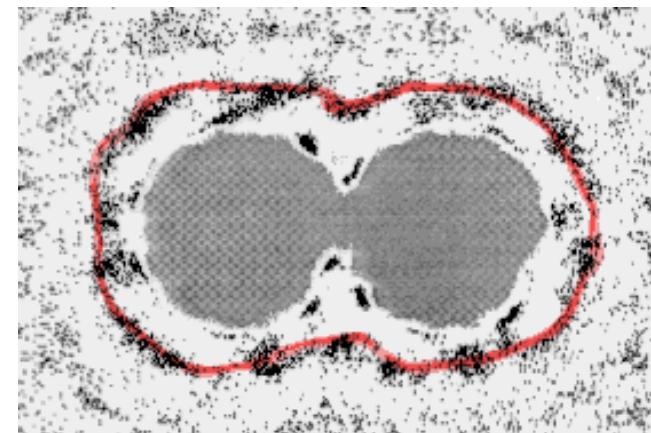
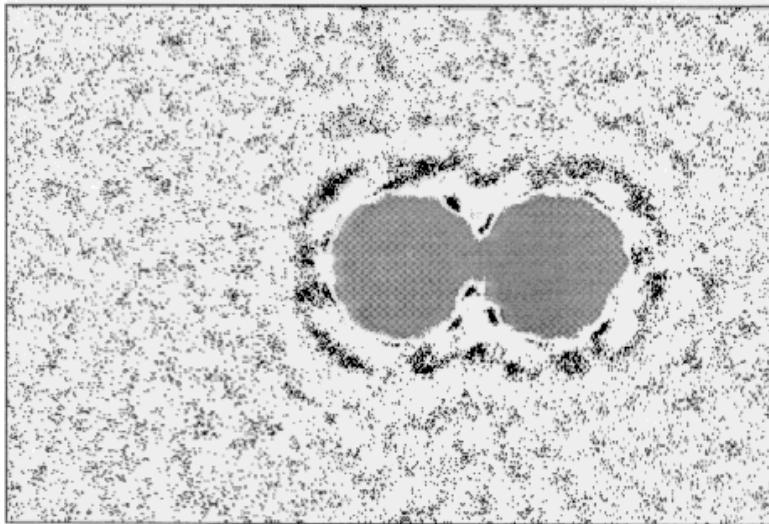
# Water VS. Ar (Helical Project- ions)

LJ & ZU

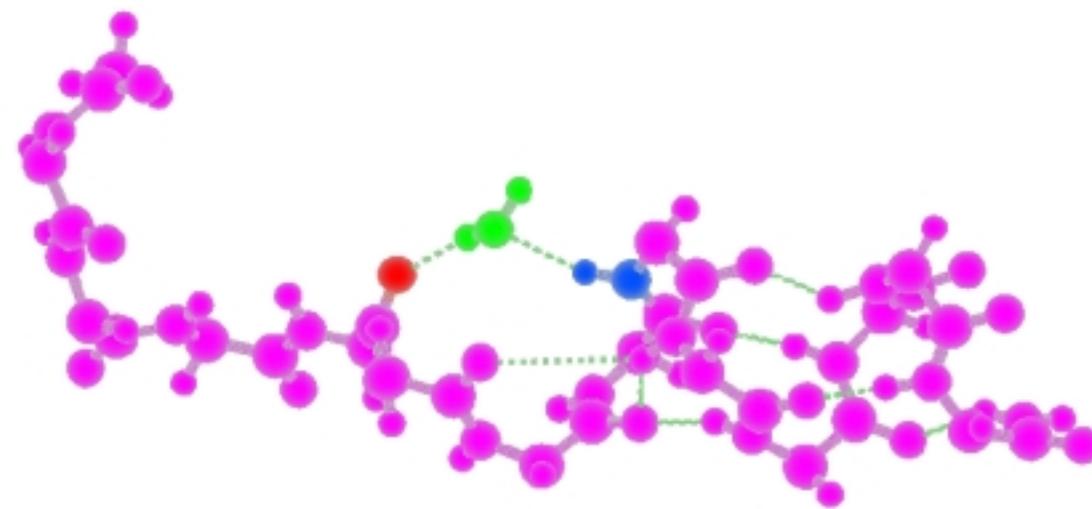


# Hydration Surface

- Bring together two helices
  - ◊ Unusually low water density in grooves and crevices — especially, as compared to uncharged water
  - ◊ Fit line through second shell



# Water Participates in Protein Unfolding



A PROTEIN HELIX CAN UNFOLD more easily in solution (than in vacuum) because water molecules can replace its helical hydrogen bonds. An unfolding helix is shown. The bottom half the helix is intact and has its helical hydrogen bonds while the top half is unfolded. In the middle a water molecule (*green*) is shown bridging between two atoms that would be hydrogen-bonded in a folded helix: the carbonyl oxygen (*red*) and the amide nitrogen (*blue*).

# Simplified Simulation

# Simplification

## BASIS OF SIMPLIFICATION

### Computational

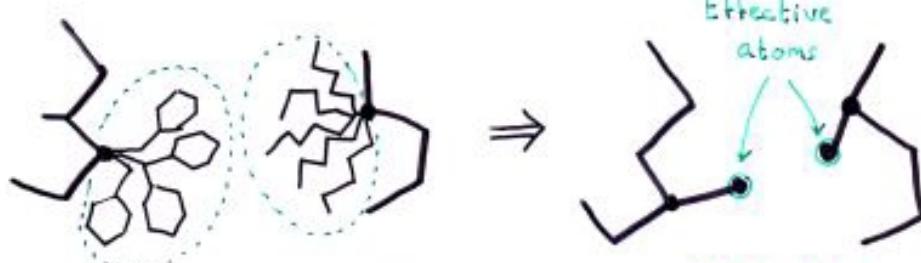
- Fewer degrees of freedom.  
Smaller space to search.
- Energy surface has less features.

Smooth surface is searched easily.



### Physical

- Time-average forces.  
Mean field.



Side chains are in continual motion and their exact position is uncertain

Illustration from M Levitt,  
Stanford University

# Simplified Protein: Lattice Models

- Cubic Lattice
- Tetrahedral Lattice

Illustration from M Levitt,  
Stanford University

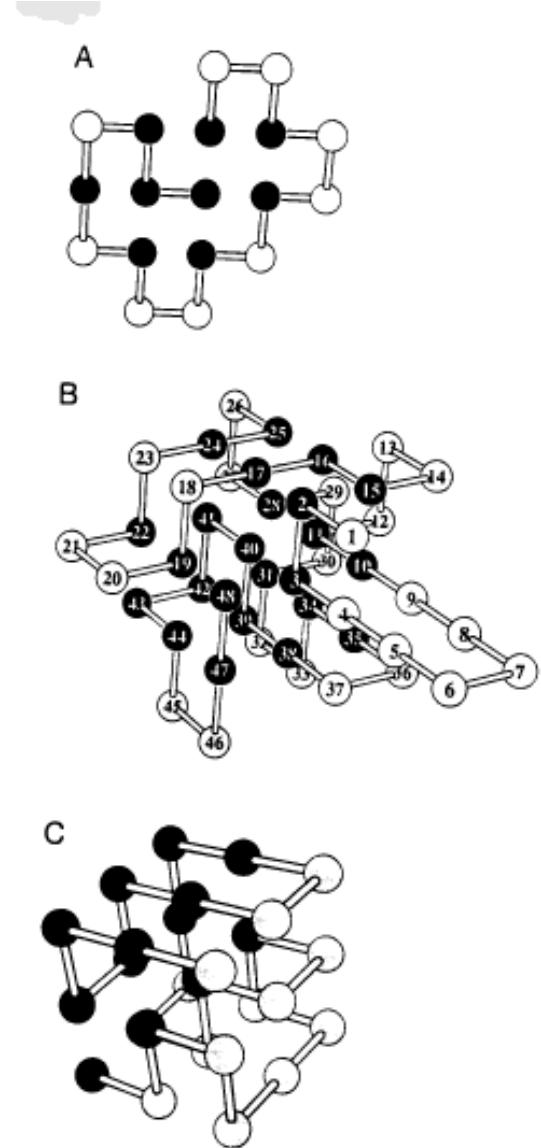
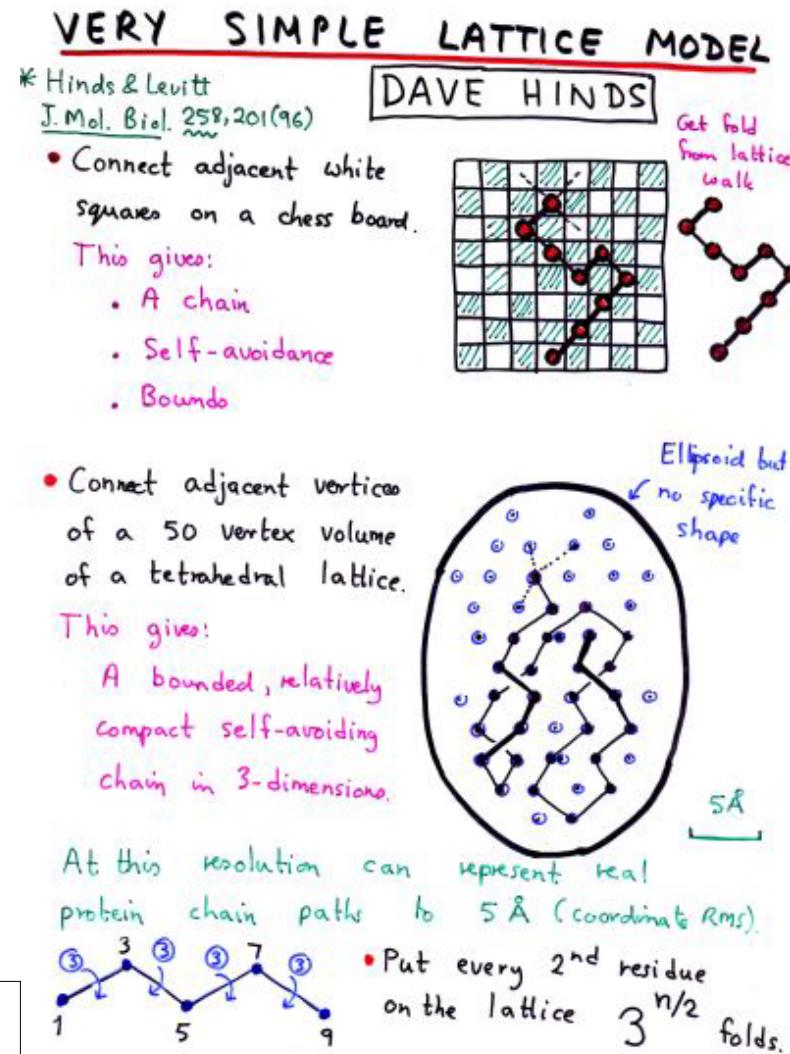
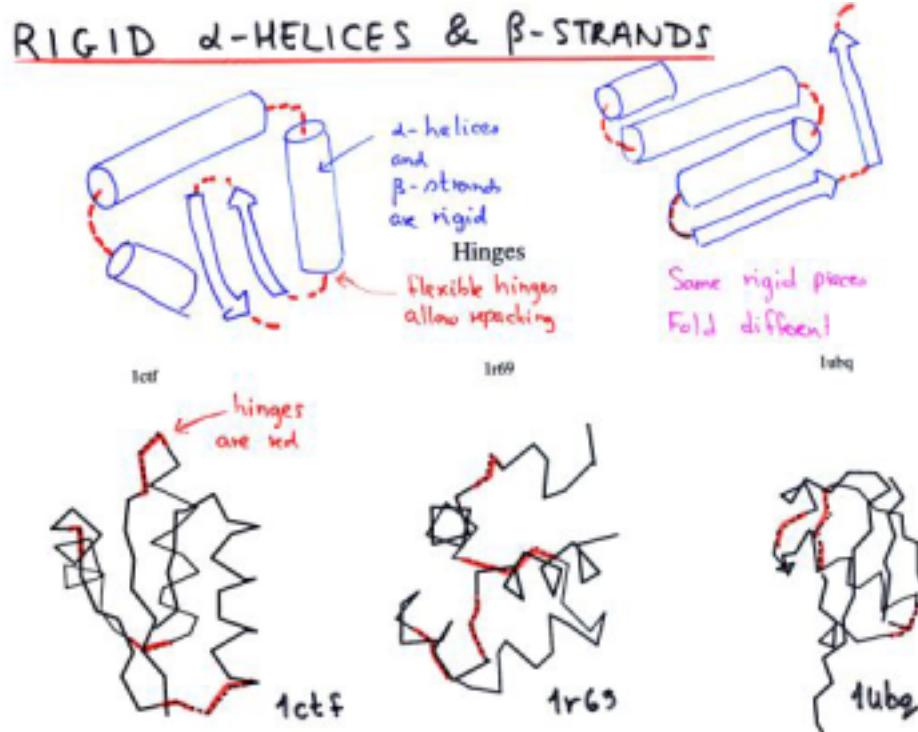
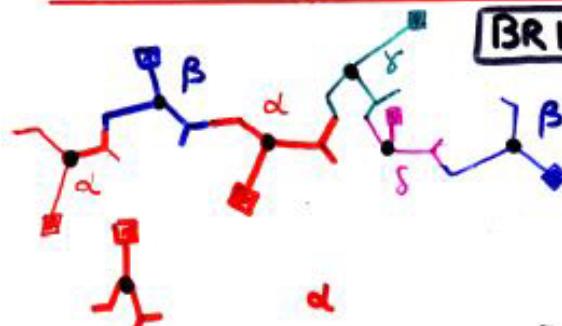


Illustration from  
Dill et al. (1990)

# Off-lattice Discrete State Models

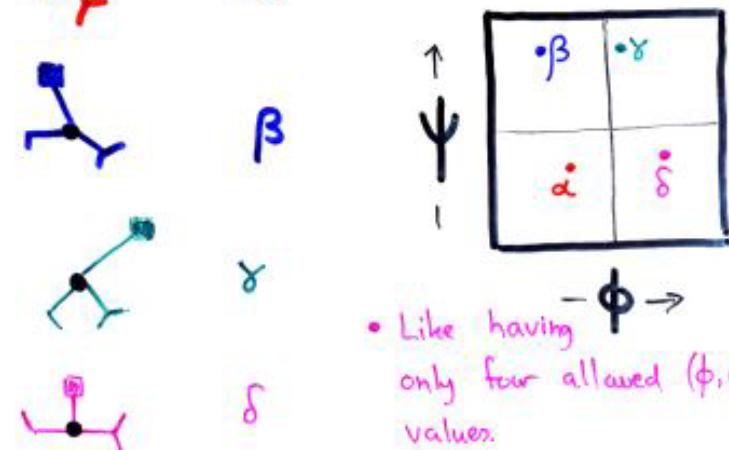


## FOUR-STATE OFF-LATTICE MODEL



BRITT PARK

- Have four rigid peptide components for all amino acids (same four for each)



- Like having only four allowed  $(\phi, \psi)$  values.
- Fit x-ray (black) with four state model (use depth limited search) Must be continuous

Illustration from M Levitt, Stanford University

# How Well Do Lattice Structures Match Real Protein Structure?

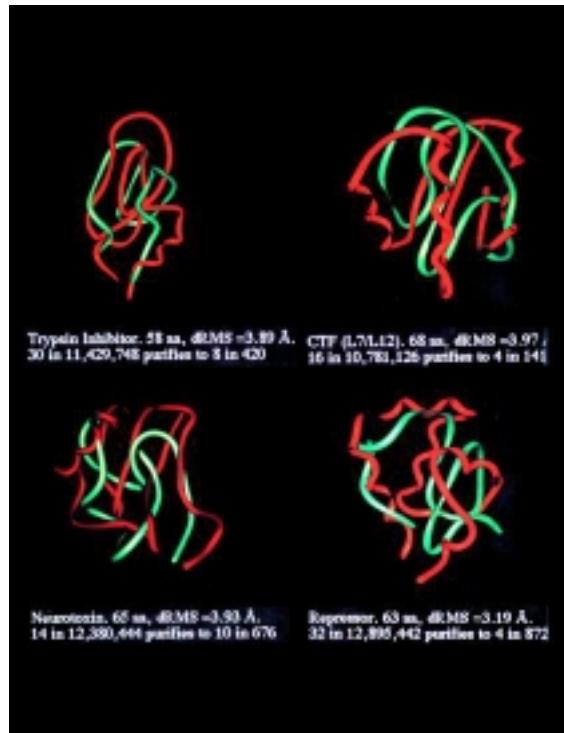


Illustration Credit: Dill et al. (1995)

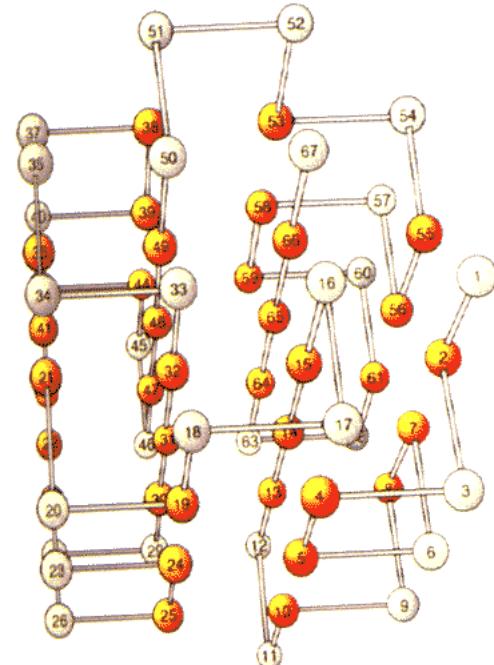
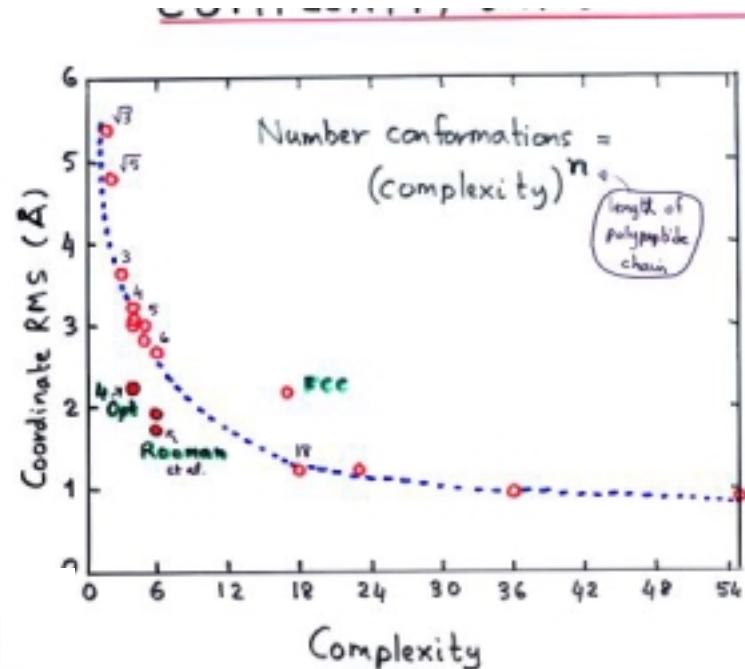
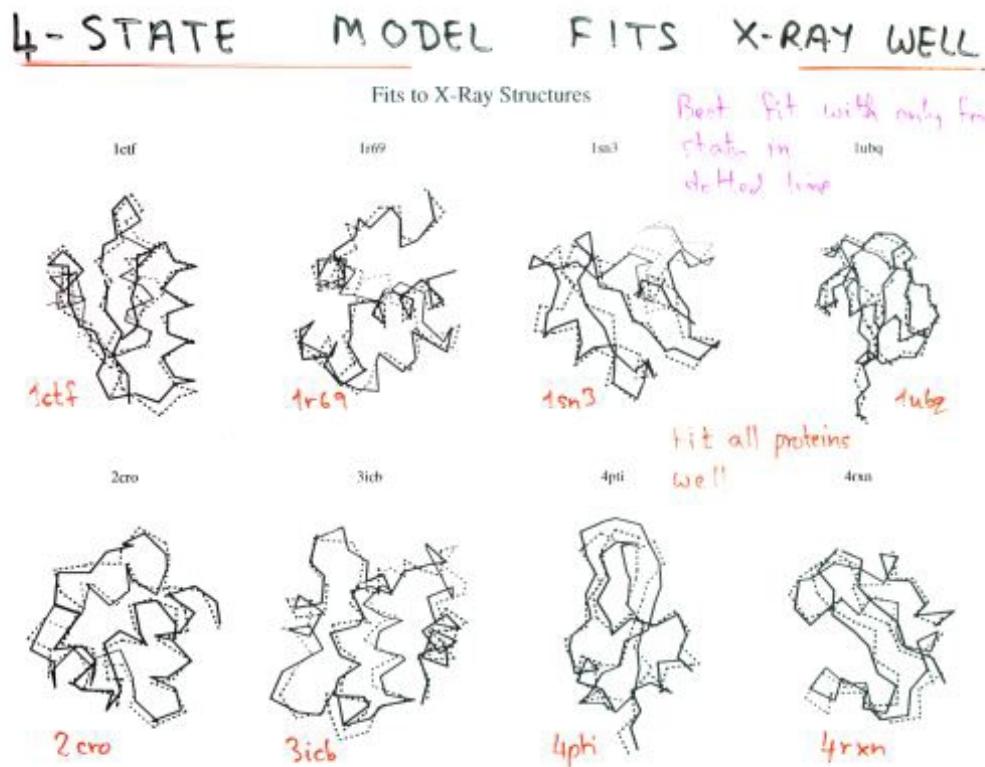


Illustration Credit: Hinds & Levitt (1992)

# How well does the off-lattice model fit?

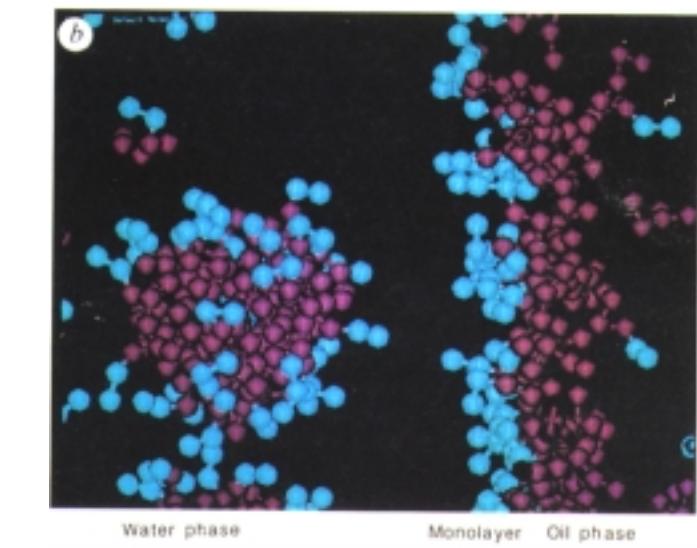
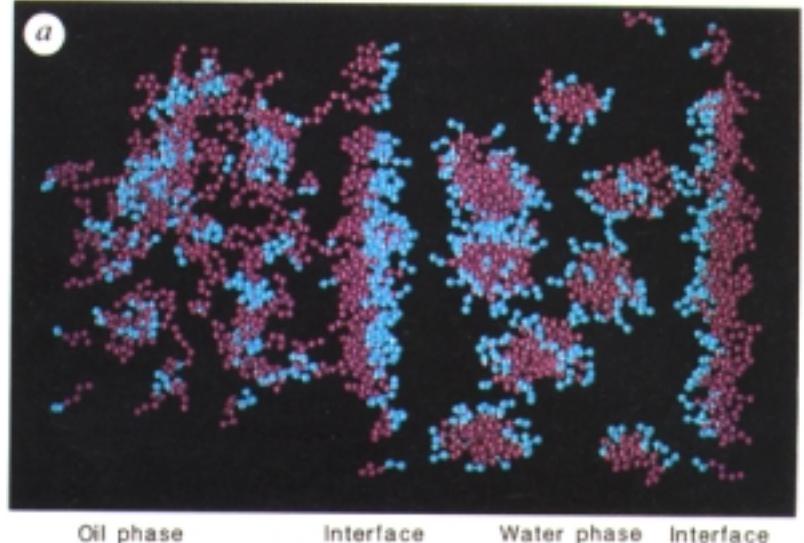


Model Complexity vs Fit to Reality

Illustration from M Levitt,  
Stanford University

# Simplified Solvent

- Smit et al. (1990) Surfactant simulation
- Three types of particles, o, w and s
  - ◊ s consists of  
 $w-w-o-o-o-o$
  - ◊ s has additional springs
- all particles interact through L-J potential
  - ◊ o-w interaction truncated so purely repulsive
- Above sufficient to give rise to the formation of micelles, membranes, &c



Figures from Smit et al. (1990)

# Review -- Basic Forces

- Basic Forces
  - ◊ Springs --> Bonds
  - ◊ Electrical
    - dipoles and induced dipoles --> VDW force --> Packing
    - unpaired charges --> Electrostatics --> charge-charge
- Electrostatics
  - ◊ All described the PBE
  - ◊  $kqQ/r$  -- the simplest case for point charges
    - Multipoles for more complex dist.
    - Validity of monopole or dipole Apx. (helix dipole?)
  - ◊ Polarization (epsilon)
    - Qualitative understanding of what it does
    - 80 vs 3

# Review -- Simulation

- Moving on an Energy Landscape
  - ◊ Minimization -- steepest descent
  - ◊ Monte Carlo
  - ◊ Molecular Dynamics
    - Know how an atom will move
  - ◊ The problems
    - Too complex --> Simplified Models
    - Potential Problems
- Analysis
  - ◊ Number density --> RDF, structural quantities
  - ◊ Dynamic quantities, correlation functions, diffusion
    - time course of variables
  - ◊ Hydrophobicity arises naturally in water simulation
    - clathrate formation
    - high heat capacity, volume effects, &c.

# Demos

- Minimization Demo
  - ◊ <http://www.javasoft.com/applets/jdk/1.0/demo/GraphLayout/example2.html>
- Adiabatic Mapping Demo
  - ◊ Molecular Motions Database
  - ◊ <http://bioinfo.mbb.yale.edu/MolMovDB>
- Rotation Matrices, Rigid Body Motion Demo
  - ◊ 1swm, 2hbs, rasmol

# References

- Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Clarendon Press, Oxford
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**, 187-217.
- Daggett, V. & Levitt, M. (1993). Realistic Simulations of Native-Protein Dynamics in Solution and Beyond. *Ann. Rev. Biophys. Biomol. Struct.* **22**, 353-380.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4**, 561-602.
- Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution *Science* **282**, 740-4.
- Eisenberg, D. & Kauzmann, W. (1969). *The Structure and Properties of Water*. Clarendon Press, Oxford.
- Franks, F. (Ed.) (1973). *Water: A Comprehensive Treatise*. New York: Plenum Press.
- Franks, F. (1983). *Water*. The Royal Society of Chemistry, London.
- Gelin, B. R. & Karplus, M. (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* **18**, 1256-1268.
- Gerstein, M. & Chothia, C. (1996). Packing at the Protein-Water Interface. *Proc. Natl. Acad. Sci. USA* **93**, 10167-10172.
- Gerstein, M. & Levitt, M. (1998). Simulating Water and the Molecules of Life. *Sci. Am.* **279**, 100-105.
- Gerstein, M. & Lynden-Bell, R. M. (1993a). Simulation of Water around a Model Protein Helix. 2. The Relative Contributions of Packing, Hydrophobicity, and Hydrogen Bonding. *J. Phys. Chem.* **97**, 2991-2999.
- Gerstein, M. & Lynden-Bell, R. M. (1993b). What is the natural boundary for a protein in solution? *J. Mol. Biol.* **230**, 641-650.
- Gerstein, M., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface: Calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**, 955-966.
- Hinds, D. A. & Levitt, M. (1992). A lattice model for protein structure prediction at low resolution. *Proc Natl Acad Sci U S A* **89**, 2536-40.

- Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144-9.
- Karplus, M. & McCammon, J. A. (1986). The dynamics of proteins. *Sci. Am.* **254**, 42-51.
- Karplus, M. & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* **347**, 631-639.
- Levitt, M. (1982). Protein conformation, dynamics, and folding by computer simulation. *Ann. Rev. Biophys. Bioeng.* **11**, 251-271.
- Levitt, M. (1983a). Molecular dynamics of a native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **168**, 595.
- Levitt, M. (1983b). Molecular dynamics of a native protein. II. Analysis and Nature of the Motion. *J. Mol. Biol.* **168**, 621-657.
- Levitt, M., Hirschberg, M., Sharon, R. & Daggett, V. (1995). Potential Energy Function and Parameters for Simulations of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. *Computer Phys. Comm.* **91**, 215-231.
- Levitt, M. & Sharon, R. (1988). Accurate Simulation of Protein Dynamics in Solution. *Proc. Natl. Acad. Sci. USA* **85**, 7557-7561.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*. Cambridge UP,
- Park, B. H. & Levitt, M. (1995). The complexity and accuracy of discrete state models of protein structure. *J Mol Biol* **249**, 493-507.

## References 2

- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes in C*. Second. Cambridge University Press, Cambridge.
- Pollack, A. (1998). Drug Testers Turn to'Virtual Patients' as Guinea Pigs. *New York Times*. Nov. 10,
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes in C*. Second. Cambridge University Press, Cambridge.
- Sharp, K. (1999). Electrostatic Interactions in Proteins. In *International Tables for Crystallography*, International Union of Crystallography, Chester, UK.
- Sharp, K. A. & Honig, B. (1990). Electrostatic interactions in macromolecules. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301-32
- Smit, B., Hilbers, P. A. J., Esselink, K., Ruppert, L. A. M., Os, N. M. v. & Schlijper (1990). Computer simulation of a water/oil interface in the presence of micelles. *Nature* **348**, 624-625.

# References

- Ille, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Clarendon Press, Oxford.
- Biosym (1994). *Discover 2.9.5 Manual*. Biosym Inc., San Diego, CA.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**, 187-217.
- Daggett, V. & Levitt, M. (1993). Realistic Simulations of Native-Protein Dynamics in Solution and Beyond. *Ann. Rev. Biophys. Biomol. Struct.* **22**, 353-380.
- Gelin, B. R. & Karplus, M. (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* **18**, 1256-1268.
- Goldstein, H. (1980). *Classical Mechanics*. 2nd edition. Addison-Wesley, New York.
- Jackson, J. (1975). *Classical Electrodynamics*. Wiley, New York.
- Karplus, M. & McCammon, J. A. (1986). The dynamics of proteins. *Sci. Am.* **254**, 42-51.
- Karplus, M. & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* **347**, 631-639.
- Levitt, M. (1982). Protein conformation, dynamics, and folding by computer simulation. *Ann. Rev. Biophys. Bioeng.* **11**, 251-271.
- Levitt, M. (1983a). Molecular dynamics of a native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **168**, 595.
- Levitt, M. (1983b). Molecular dynamics of a native protein. II. Analysis and Nature of the Motion. *J. Mol. Biol.* **168**, 621-657.
- Levitt, M. (1983c). Protein folding by restrained energy minimization and molecular dynamics. *J Mol Biol* **170**, 723-64.
- Levitt, M., Hirschberg, M., Sharon, R. & Daggett, V. (1995). Potential Energy Function and Parameters for Simulations of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. *Computer Phys. Comm.* **91**, 215-231.

# References 2

- Levitt, M. & Sharon, R. (1988). Accurate Simulation of Protein Dynamics in Solution. *Proc. Natl. Acad. Sci. USA* **85**, 7557-7561.
- Marion, J. B. & Heald, M. A. (1980). *Classical Electromagnetic Radiation*. Academic Press, New York.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*. Cambridge UP,
- Pettitt, B. M. & Karplus, M. (1985). The Potential of Mean Force Surface for the Alanine Dipeptide in Aqueous Solution: A Theoretical Approach. *Chem. Phys. Lett.* **121**, 194-201.
- Atkins, P. (1990). *Physical Chemistry*. Oxford UP
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes in C*. Second. Cambridge University Press, Cambridge.
- Purcell, E. M. (1985). *Electricity and Magnetism*. McGraw-Hill, New York.
- Brünger, A. T. (1993). *X-PLOR 3.1, A System for X-ray Crystallography and NMR*. Yale University Press, New Haven.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-60.
- Rice, L. M. & Brunger, A. T. (1996). Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19**, 277-290