

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

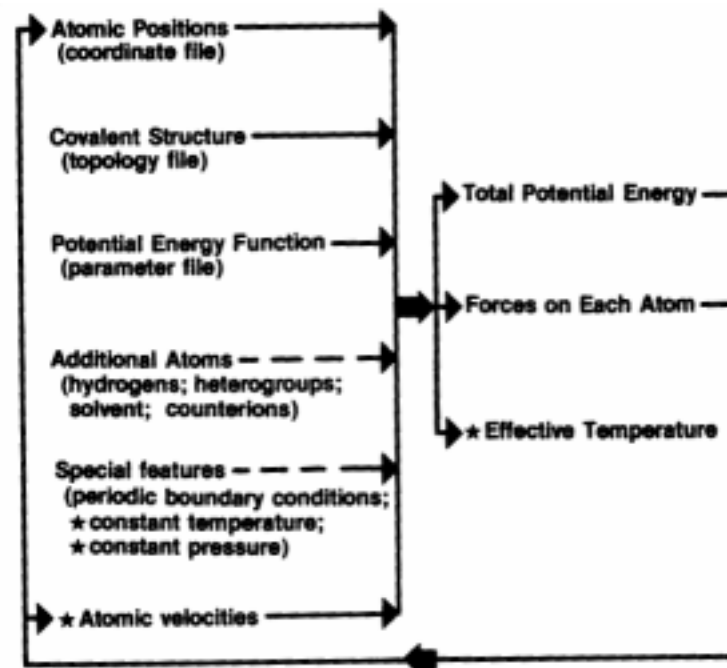


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of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

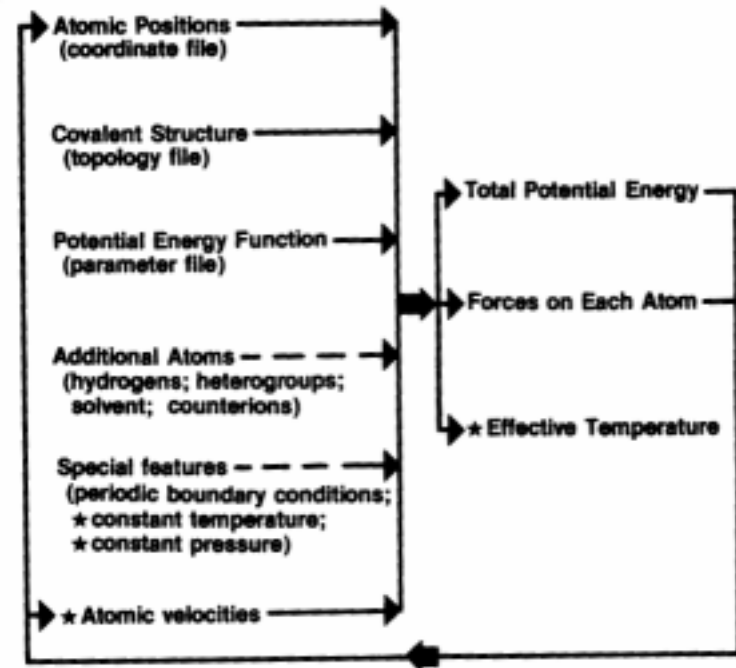


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of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

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 i th atom
in the molecule at
timestep t
- Rigid-body Translation of
all i atoms

- ◊ For each atom atom i do
 $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}$

- 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 ϕ, θ, ψ (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

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

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

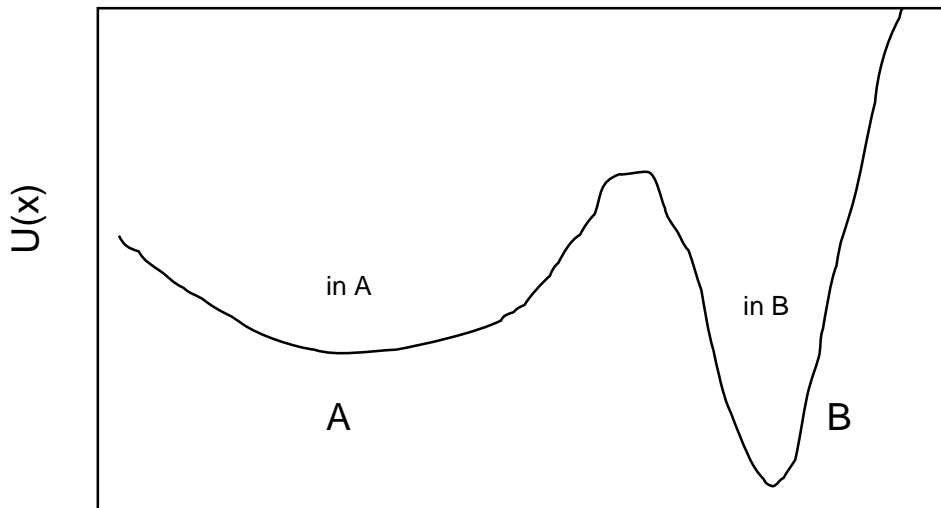
Energy and Entropy

- 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

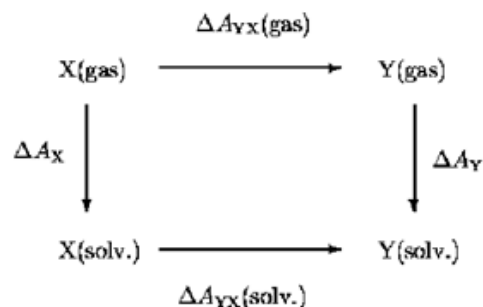
- Relative Prob

- ◇ $P(A)/P(B) = \frac{\exp(-G(A)/kT)}{\exp(-G(B)/kT)}$

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:



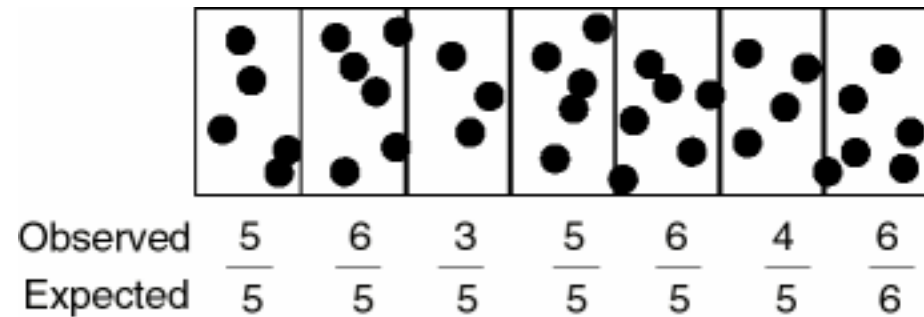
where $\Delta\mu_X$ and $\Delta\mu_Y$ are, respectively, the free energy of solvation of X and Y, and $\Delta\mu_{YX}(\text{gas})$ and $\Delta\mu_{YX}(\text{solv.})$ are the free energies of mutating X in Y in, respectively, in the gas phase and the solution phase.
(*Computational alchemy.*)

The differences of free energies of solvation is

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

Text block adapted
from on-line notes
at Rutgers
Chemistry

Number Density

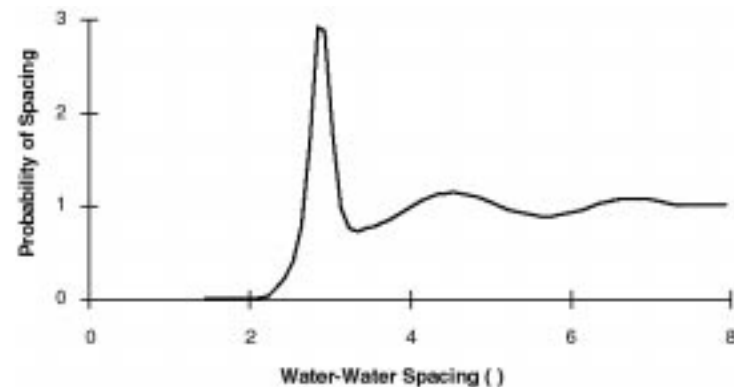


= Number of atoms per unit volume averaged over simulation divided by the number you expect to have in the same volume of an ideal “gas”

Spatially average over all directions gives

1D RDF =

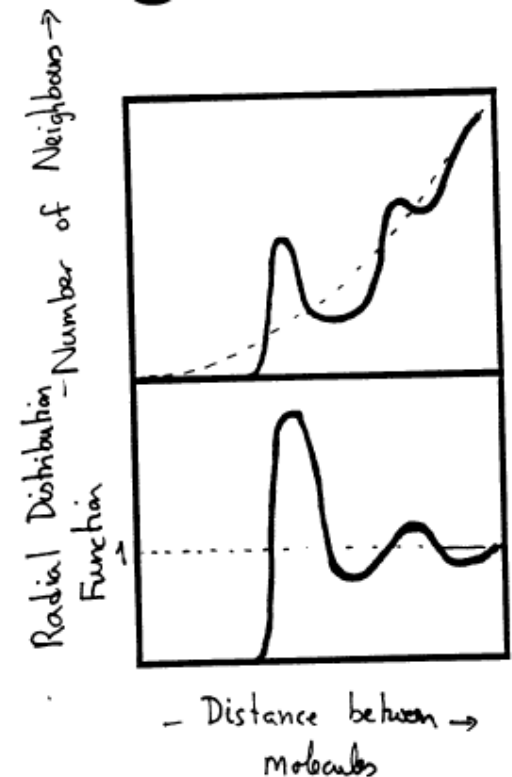
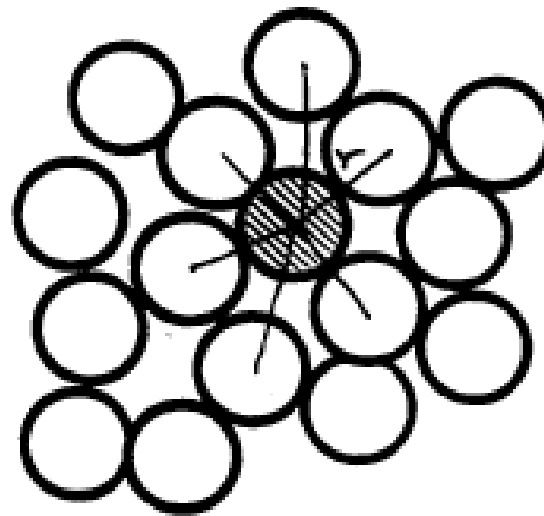
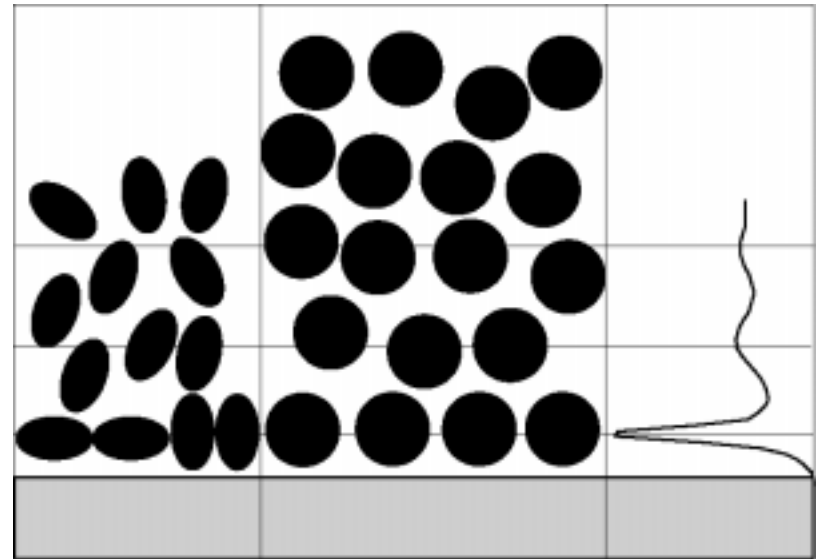
$$\frac{[\text{Avg. Num. Neighbors at } r]}{[\text{Expected Num. Neighbors at } r]}$$



“at r ” means contained in a thin shell of thickness dr and radius r_{13}

Number Density (cont)

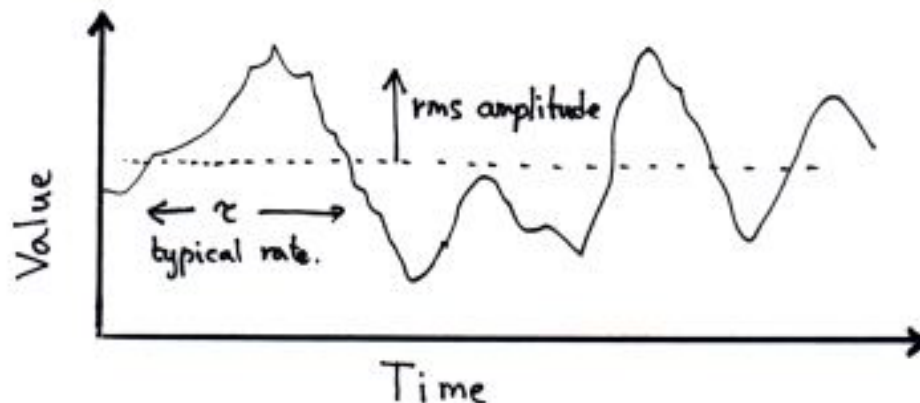
- Advantages: Intuitive,
Relates to scattering expts
- D/A: Not applicable to real proteins
 - ◇ 1D RDF not structural
 - ◇ 2D proj. only useful with "toy" systems
- Number densities measure spatial correlations, not packing
 - ◇ Low value does not imply cavities
 - ◇ Complicated by asymmetric molecules
 - ◇ 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
$$R = \sqrt{\langle (a(t) - \langle a(t) \rangle)^2 \rangle}$$
$$R = \sqrt{\langle a(t)^2 - 2a(t)\langle a(t) \rangle + \langle a(t) \rangle^2 \rangle}$$
$$R = \sqrt{\langle a(t)^2 \rangle - \langle a(t) \rangle^2}$$
 - similar to SD
 - fluctuation
- Relevant variables include bond length, solvent molecule position, H-bond angle, torsion angle

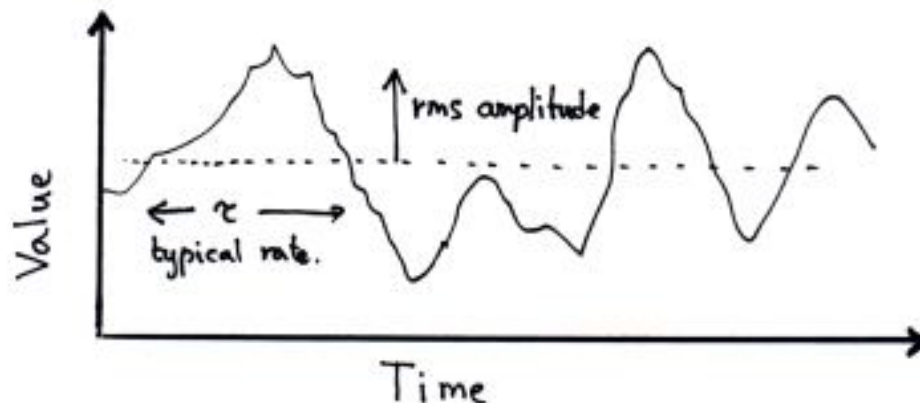
Illustration from M Levitt,
Stanford University



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

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

- 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)

Observed RMS values

COMPARISON OF OVERALL VALUES

Property	Value		
	in vacuo	in soln.	expt.
• All-Atom R.M.S. Deviation (Å)	2.60	1.55	1.3(0.5)
• C ^α Fluctuation (Å)	0.54	0.43	0.68
• Radius of Gyration (Å)	10.9	11.5	11.5

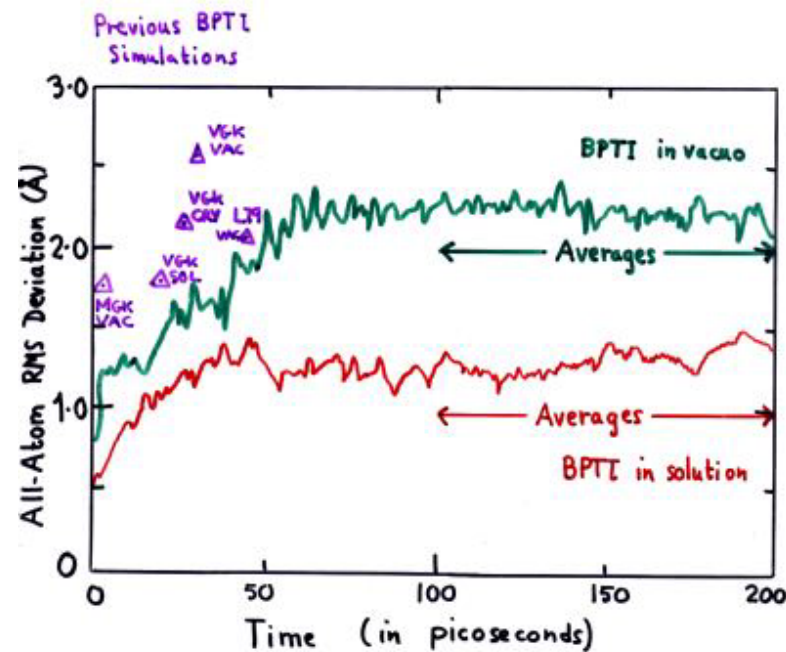


Illustration from M Levitt,
Stanford University

Other Things to Calculate

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

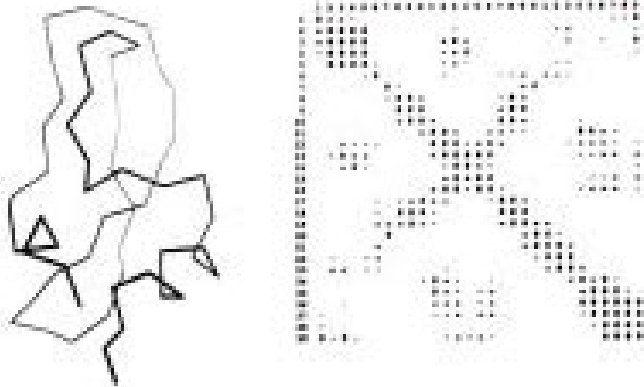
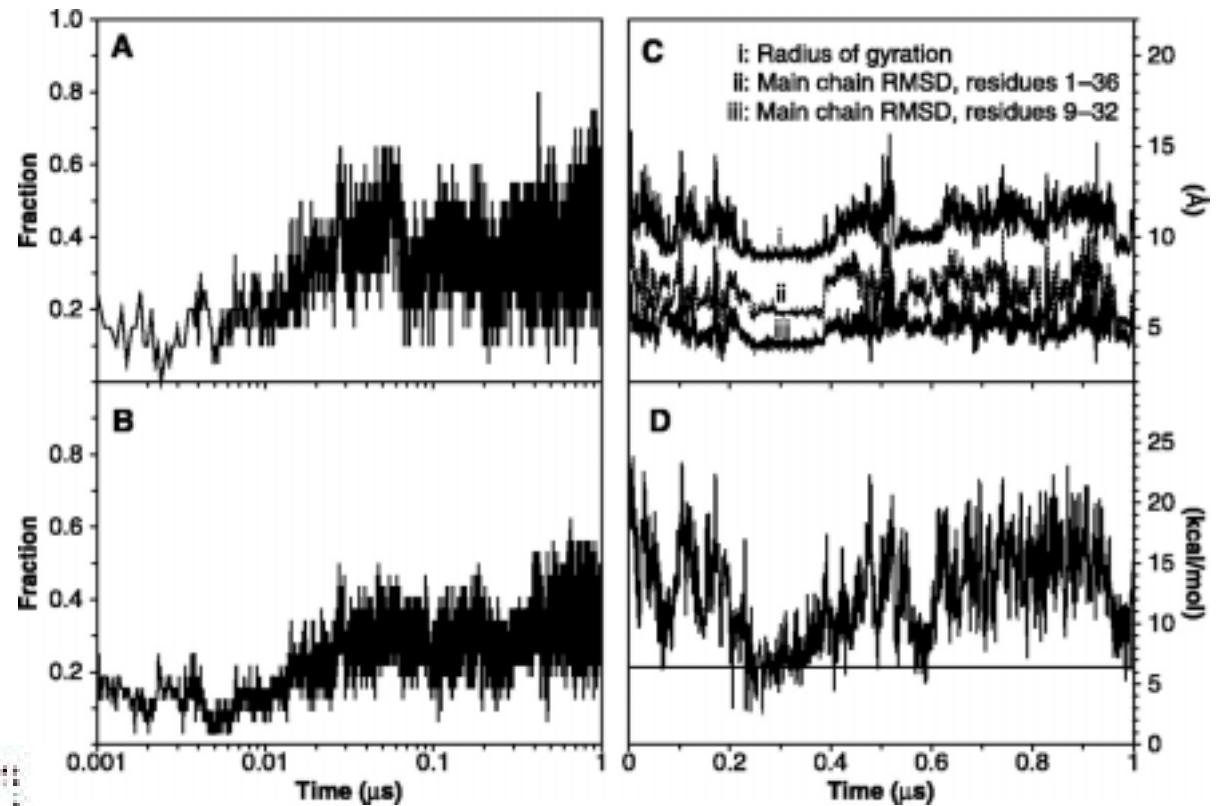


Illustration and Caption from
Duan & Kollman (1998)



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 Å², for the surface areas of nonpolar, polar, sulfur, charged oxygen, and charged nitrogen, respectively). The straight line¹⁹ represents the SFE of the native structure.

Monitor Stability of Specific Hydrogen Bonds

HYDROGEN BONDS

Secondary Structure	O..H Pair	Stability (%)	
		in vacuo	in soln.

β-hairpin

35..18	12	57
18..35	85	63
33..20	91	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

α-helix

47..51	76	66
48..52	93	90
49..53	90	98
50..54	78	90
51..55	73	93
52..56	-	42

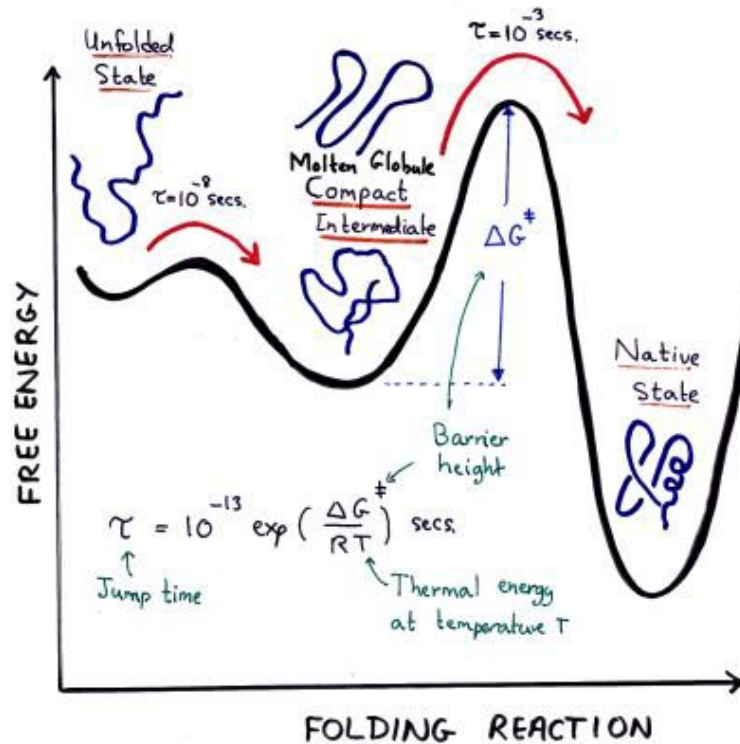
Illustration from M Levitt,
Stanford University

- 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

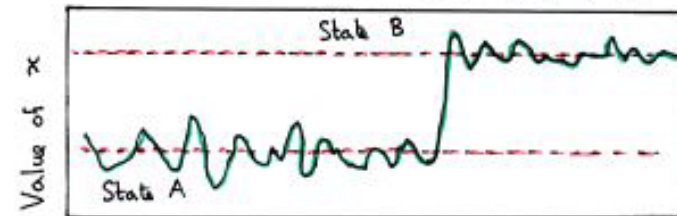
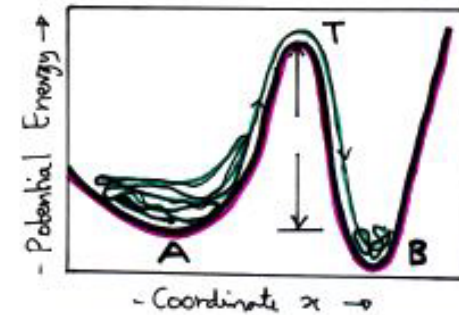
ENERGY LANDSCAPES

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 pico seconds). 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
Boltzmann constant

$$\left(\frac{h}{kT} \right) \approx 0.16 \text{ picoseconds at } T = 300^\circ K (27^\circ C).$$

(remember as $h\nu_0 = kT$, $\nu_0 = 208 \text{ cm}^{-1}$, $\tau_0 = 0.16 \text{ ps}$)

Illustrations from M Levitt, Stanford University

(c) M Gerstein (<http://bioinfo.mbb.yale.edu>)

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. Ushh!

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

dielectric const IN potential OUT ionic strength IN fixed charges 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. Ughh!

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

dielectric const IN potential OUT ionic strength IN fixed charges IN

- No moving ions, Constant Dielectric \rightarrow Poisson's Eq.

$$\nabla^2 \phi(\vec{r}) = \frac{4\pi}{kT\epsilon} \rho(\vec{r})$$

- Finite Difference Soln. to PDE

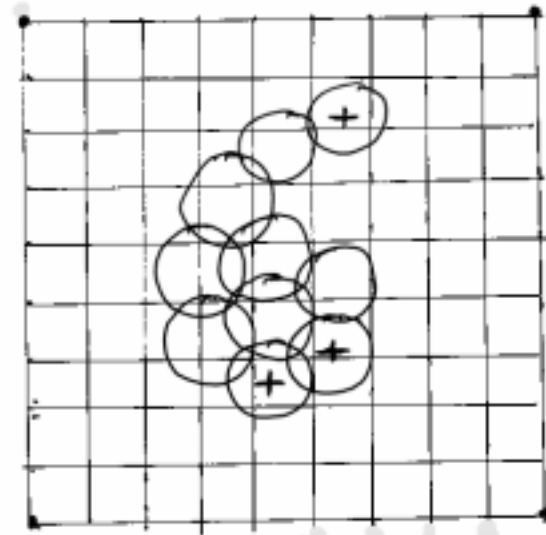
(PDE has deriv. WRT to 2 var.
ODE like Newton's eq. has deriv. WRT to 1 var.)

- $\nabla^2 \phi(\vec{r}) = \frac{4\pi}{kT\epsilon} \rho(\vec{r})$

- in 2D $\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \underbrace{C}_{\text{const.}} \rho(\vec{r})$

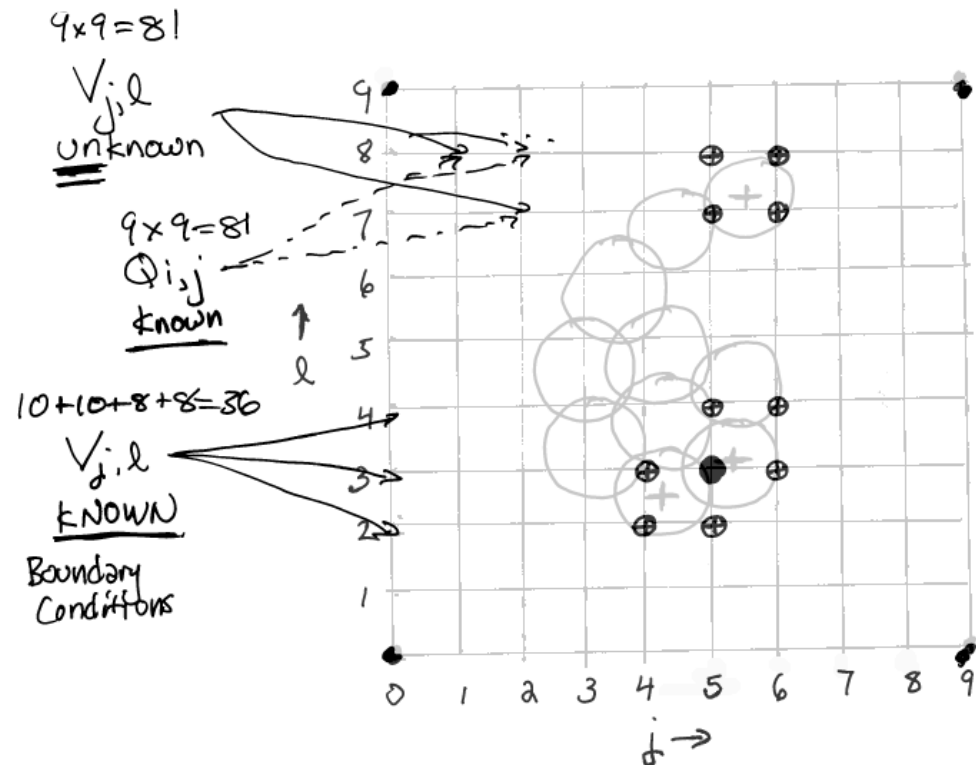
- $\frac{\partial^2 \phi}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial \phi}{\partial x} \right) = \frac{\partial}{\partial x} \left(\frac{V_{j+1} - V_j}{\Delta} \right) = \frac{(V_{j+1} - V_j) - (V_j - V_{j-1})}{\Delta^2}$
($\partial x = \Delta$, $\partial \phi = V_{j+1} - V_j$)

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



Protein on a Grid

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



$\oplus = \frac{q}{2}$
 $\bullet = \frac{-q}{2}$

Demand Consistency on the Grid

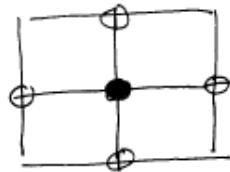
- $V_{j+1,l} + V_{j-1,l} + V_{j,l+1} + V_{j,l-1} - 4V_{j,l} = \Delta^2 C Q_{j,l}$
- System of Equations \rightarrow solve for unknown $V_{j,l}$
- Matrix Inversion in Finite Diff. method

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

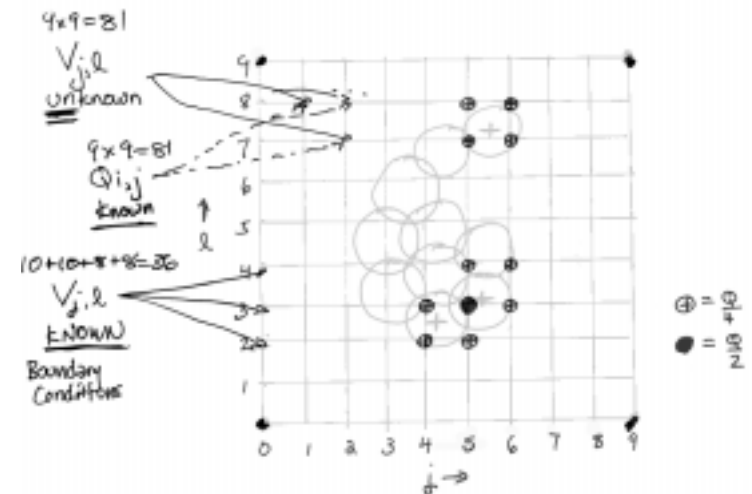
$$\nabla^2 V - 4\pi\beta = \left(\frac{\partial V}{\partial t}\right) \rightarrow 0 \text{ at } t = \infty$$

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

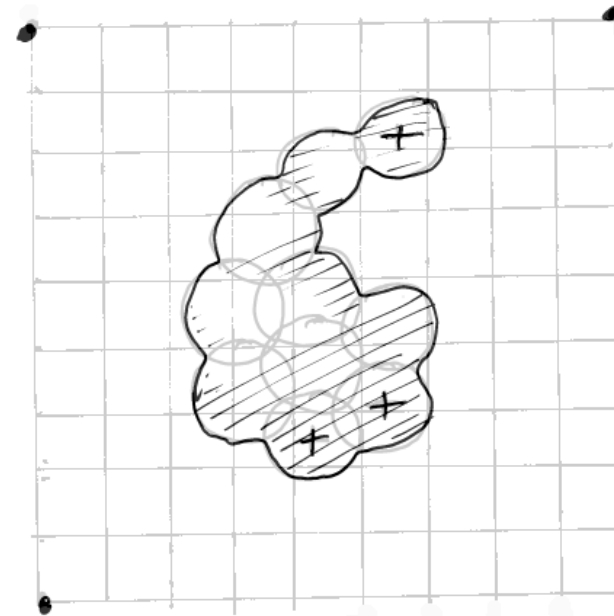
For intuition ONLY
-- Don't need to
know in detail!!



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



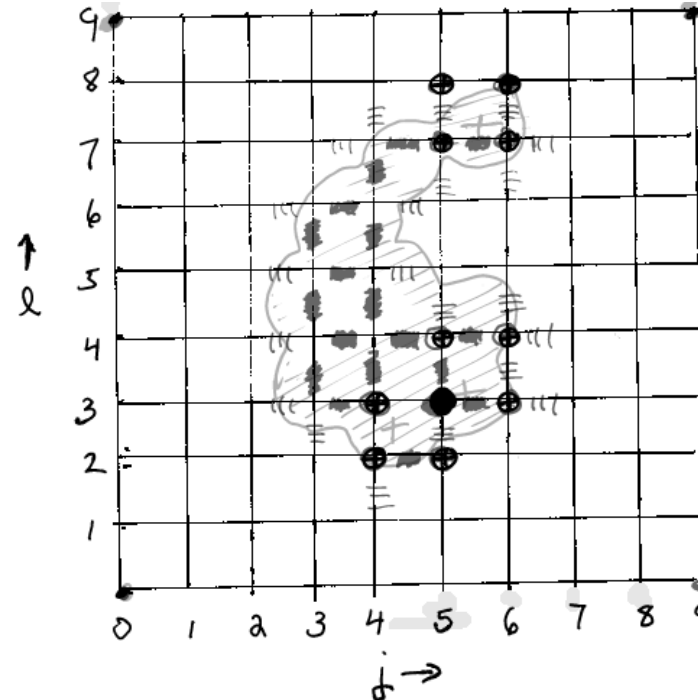
Adding a Dielectric Boundary into the Model



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

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

$$\frac{1}{\Delta^2} (\epsilon(j \rightarrow j+1) (V_{j+1} - V_j) - \epsilon(j-1 \rightarrow j) (V_j - V_{j-1}))$$



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

$\text{diagonal lines} = 2$
 $\text{||||} = 40$
 $\text{---} = 80$

$\oplus = \frac{q}{4}$
 $\bullet = \frac{q}{2}$

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.

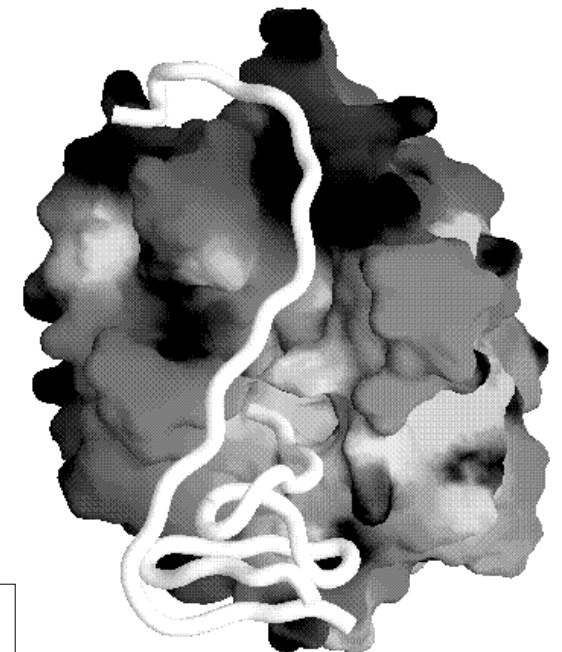
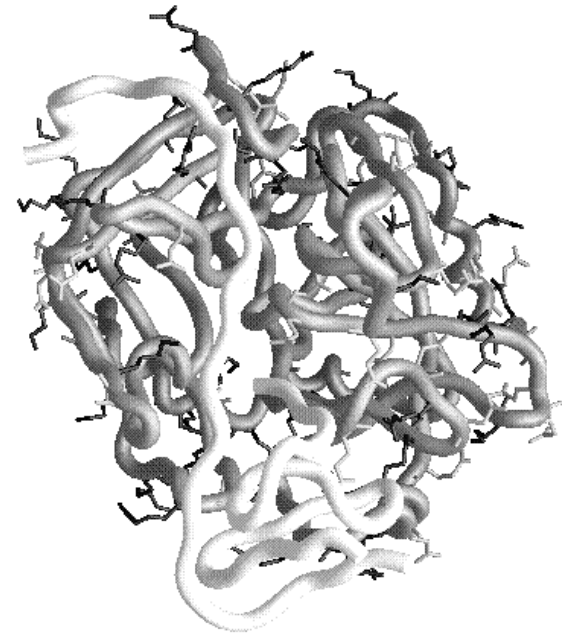
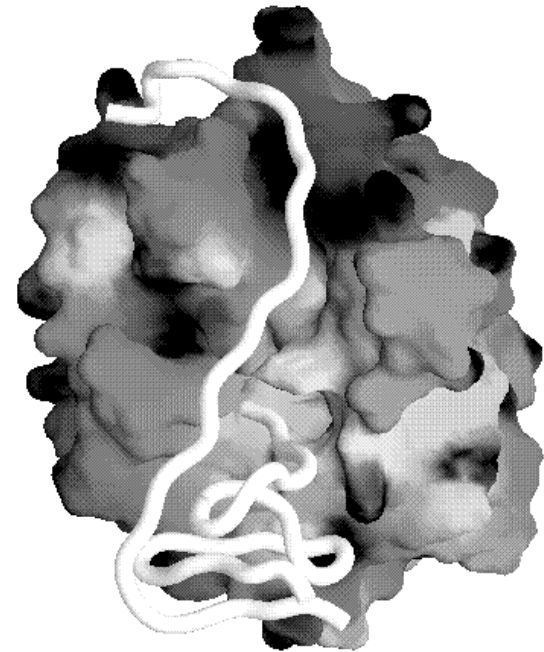


Illustration Credit: Sharp (1999)
Text captions also from Sharp (1999)

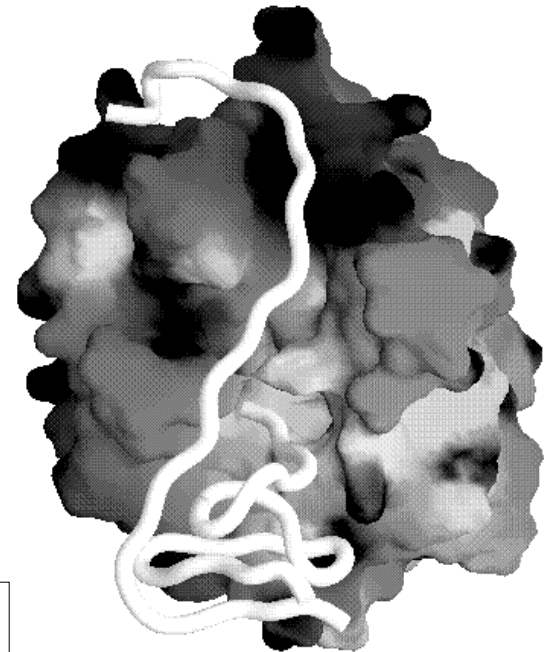
Increasing Ionic Strength

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)



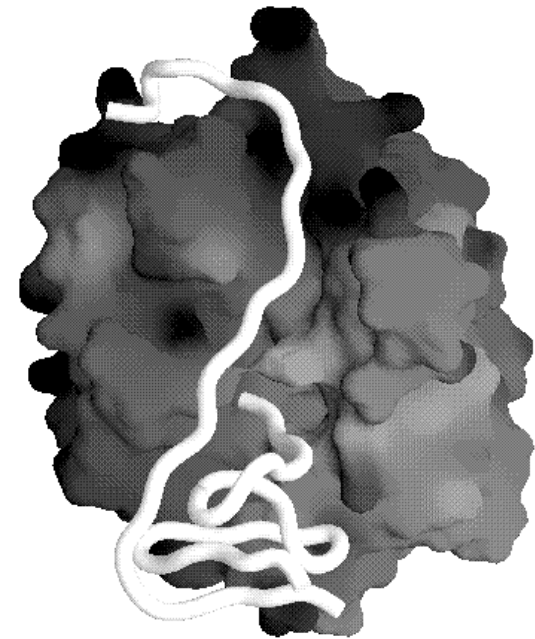
TOP shows the effect of increasing ionic strength on the potential distribution, shrinking the regions of strong potential in comparison to BOTTOM.

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.



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.

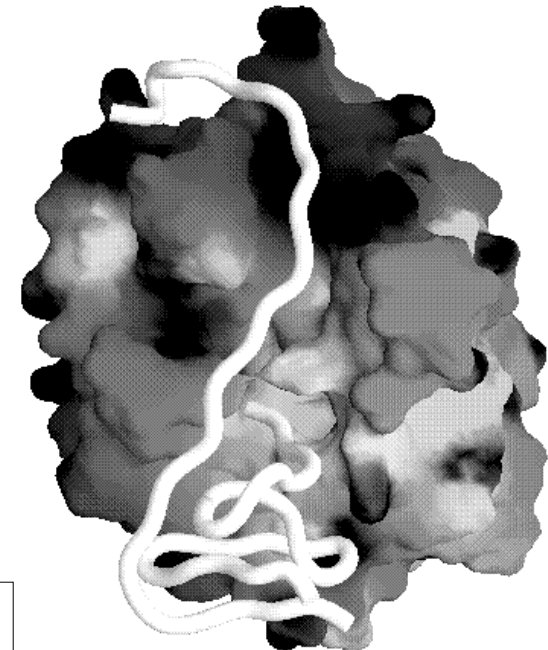


Illustration Credit: Sharp (1999)
Text captions also from Sharp (1999)

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^0_a . Where $pK^0_a = -\log_{10}(K^0_a)$, K^0_a is the dissociation constant for the reaction $H^+A = H^+ + A$ and A can be an acid or a base. The pK^0_a 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 $pK^0_a = 3.85$. 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):

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

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

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

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

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

where ϕ_i^{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:

$$\Delta\Delta G^{\text{titr}} = \sum_i \left(q_i^d \langle \phi_i \rangle_{\text{pH}, c, \Delta V}^d - q_i^p \langle \phi_i \rangle_{\text{pH}, c, \Delta V}^p \right) \quad (14)$$

3. Other Charges

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 ΔV 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 , ΔV . From this one may calculate the mean ionization state of any group as a function of pH, ΔV etc. For large N this becomes impractical, but various approximate schemes work well, including a Monte-Carlo procedure

Text block from
Sharp (1999)

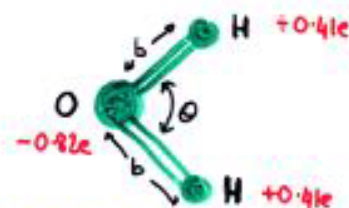
Water Simulation and Hydrophobicity

Simulating Liquid Water

SIMULATING LIQUID WATER

- Very simple model

- 3 interaction centers
- Completely flexible
- Smooth cutoff at 6 Å (list to 8 Å)

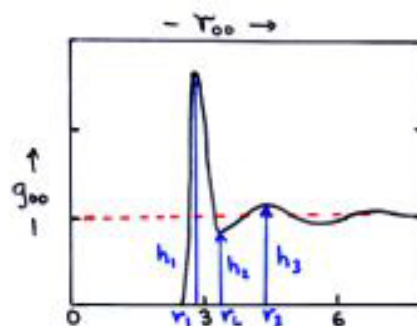


Electrostatics } long range forces
Van der Waals }

- Good fit to experiment

Property	(25 °C)	
	Experiment	Simulation
Potential energy (kcal/mol)	-9.2	-9.5
Pressure (atmospheres)	1	-61
Classical Specific Heat (cal/°K)	27	26
Diffusion Constant ($\text{\AA}^2/\text{ps}$)	0.23	0.22
Rotational Relaxation Time (ps)	2	1.6

Radial Distribution Function



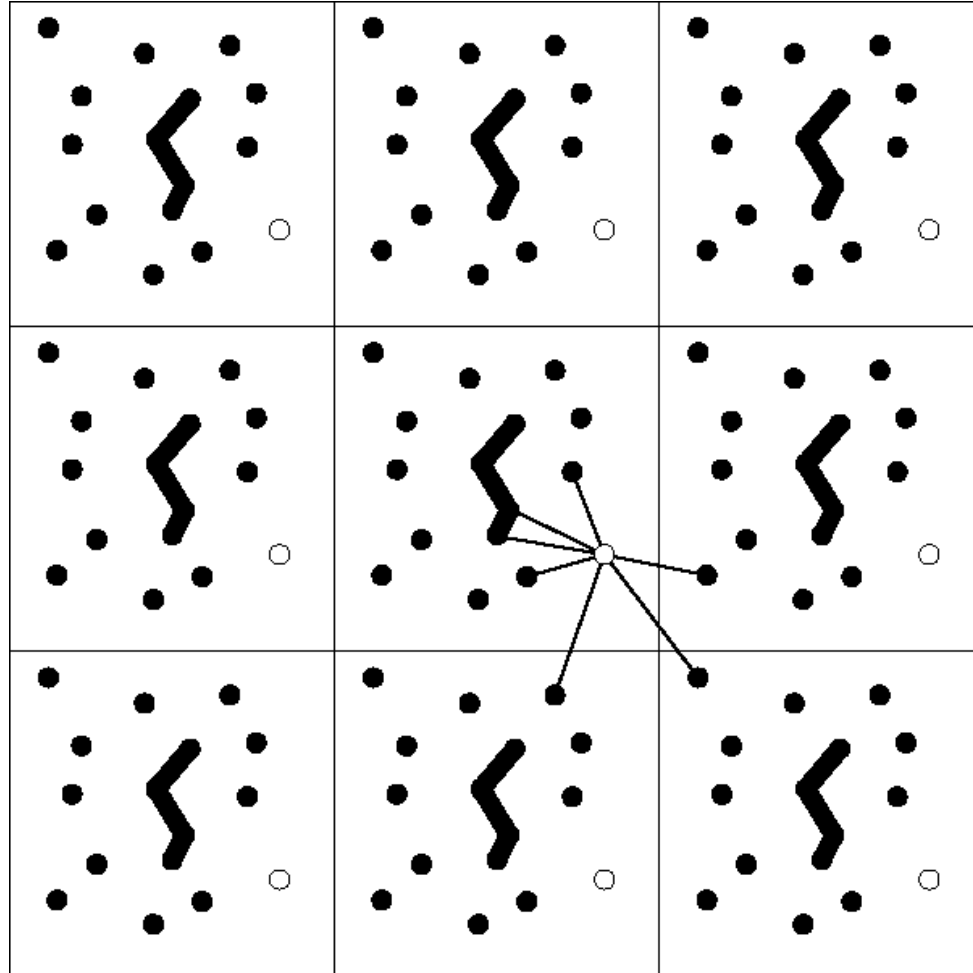
r_1	2.8	2.7
h_1	2.5 3.0*	3.2
r_2	3.3	3.3
h_2	0.8	0.8
r_3	4.6	4.3
h_3	1.11	1.09

* Calibration error fixed after 15 years of experiment

Illustrations from
M Levitt, Stanford
University

Periodic Boundary Conditions

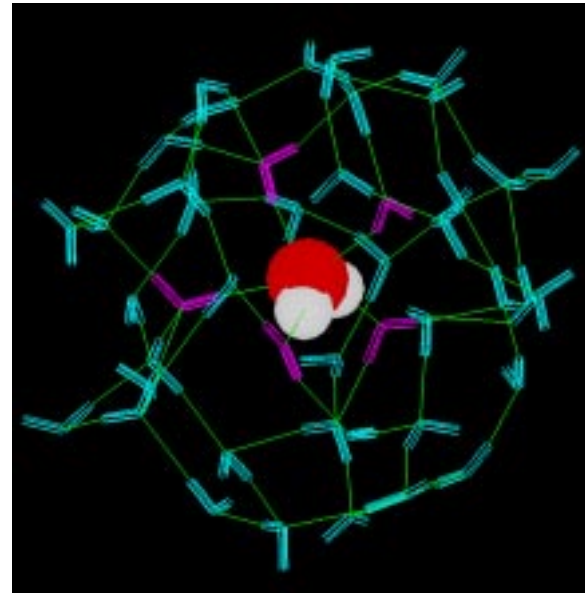
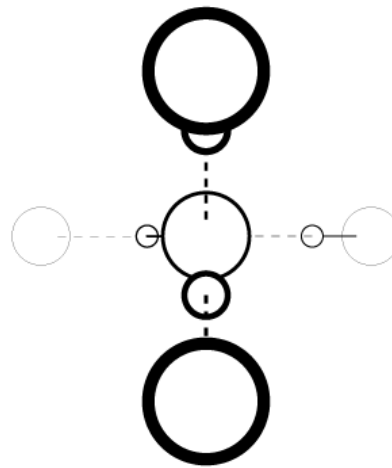
- Make simulation system seem larger than it is
- Ewald Summation for electrostatics (Fourier transform)



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 ($H1$) 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 a 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.

(c) M Gerstein (<http://bioinfo.mbb.yale.edu>)



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.

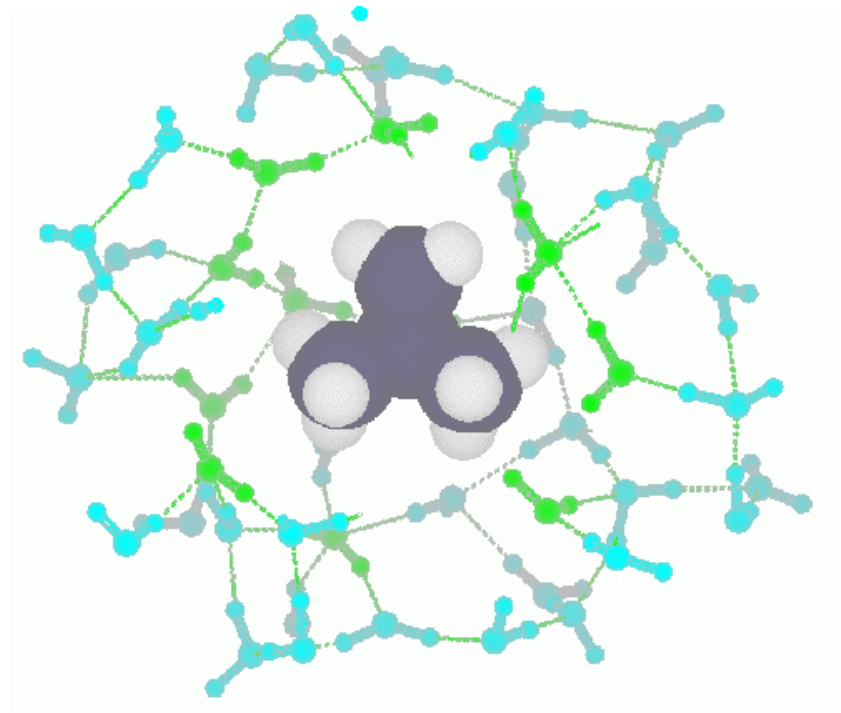
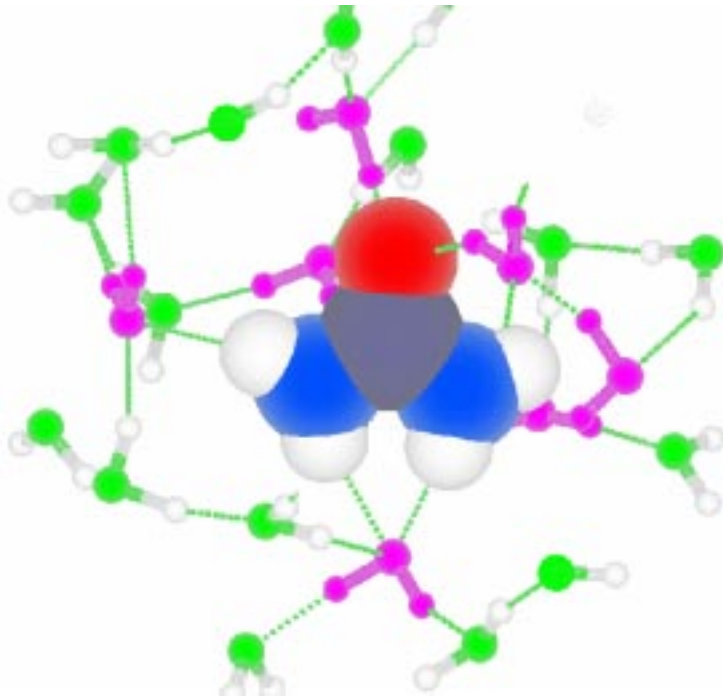
In contrast, water molecules are not able to hydrogen bond to methane, an insoluble, non-polar solute. They would rather interact with each other. The methane molecules, moreover, can favorably interact with each other through attractive dispersion forces. One can see how this situation leads to methane 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

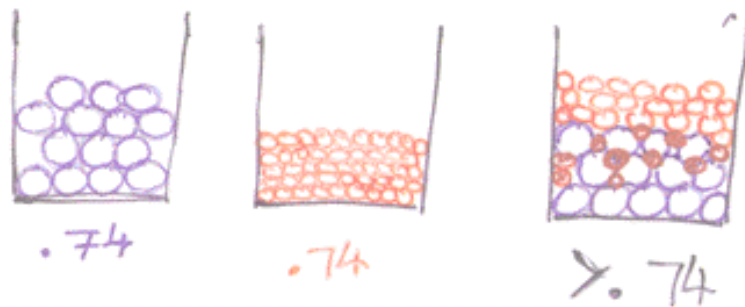
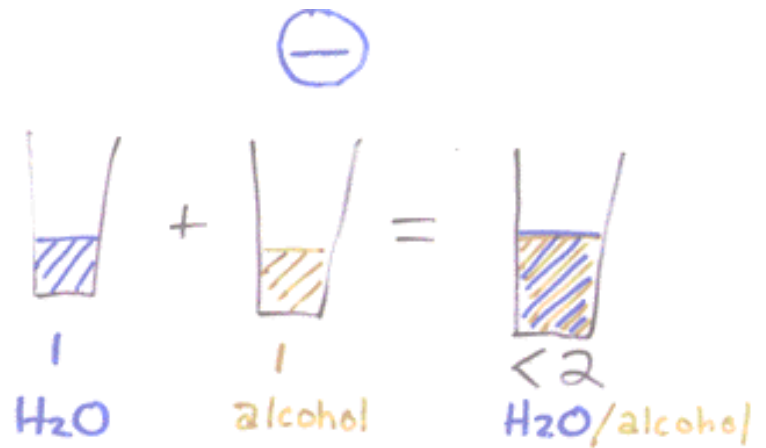


POLAR AND NON-POLAR SOLUTES have very different effects on water structure. We show two solutes that have the same Y-shaped geometry but different partial charges. The polar solute, urea (*left*), has partial charges on its atoms. Consequently, it is able hydrogen-bond to water molecules and to fit right 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

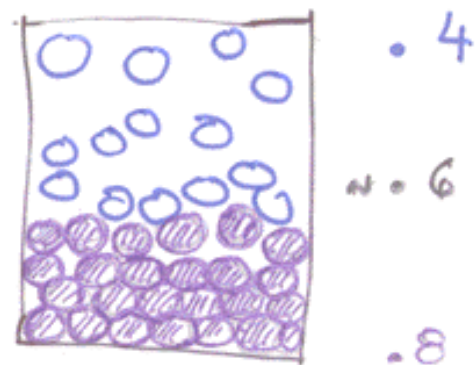
- Hydrophobic hydration is unfavorable (G) but the reason is different at different T
 - ◇ entropically (S) unfavorable at low temperatures because of ordering
 - ◇ enthalpically (H) unfavorable at high temperatures because of unsatisfied H-bonds
- Volume of mixing is negative
- Compressibility
- High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

Ways of Rationalizing Packing



⊕

TIGHT CORE
($>$ organic xtals)



Compare standard volumes with amino acids CRYSTAL volumes

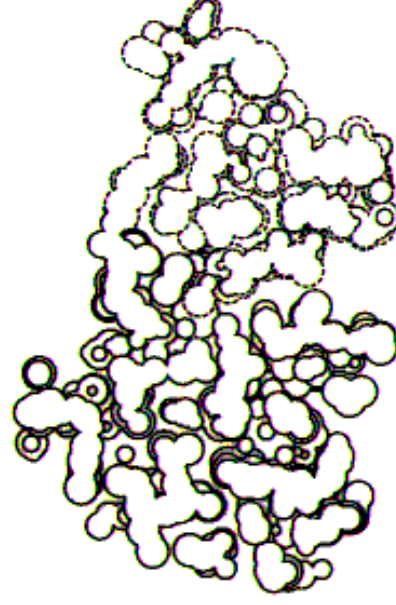
Example residue volume: Leu (\AA^3)

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

Example atomic volume: $-\text{CH}_2-$ (\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

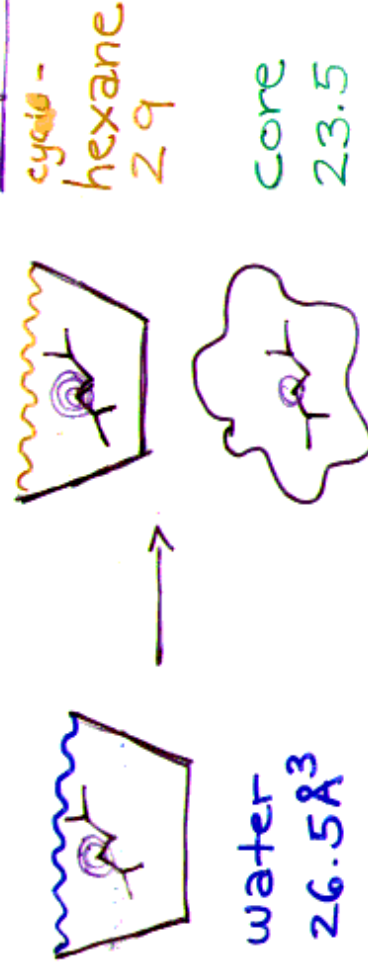
$$V_{\text{SIDECHAIN}} - V_{\text{SOLUTION}} = V_{\text{CORE}} \quad \begin{matrix} \text{SIDECHAIN} \\ \text{SOLUTION} \end{matrix} \quad \begin{matrix} \text{CORE} \\ \text{(Cohn et al. '34)} \\ \text{(Rao et al. '84)} \end{matrix}$$

= + 4 ALIPHATICS (AYLIP)

= ~ 0 POLARS, AROMATICS (MCFY)
(WST)

= - 7 CHARGED, AMIDE (HNDQ)
(ERK)

SOLUTION-TRANSFER Models Predict
Opposite Result for Aliphatics



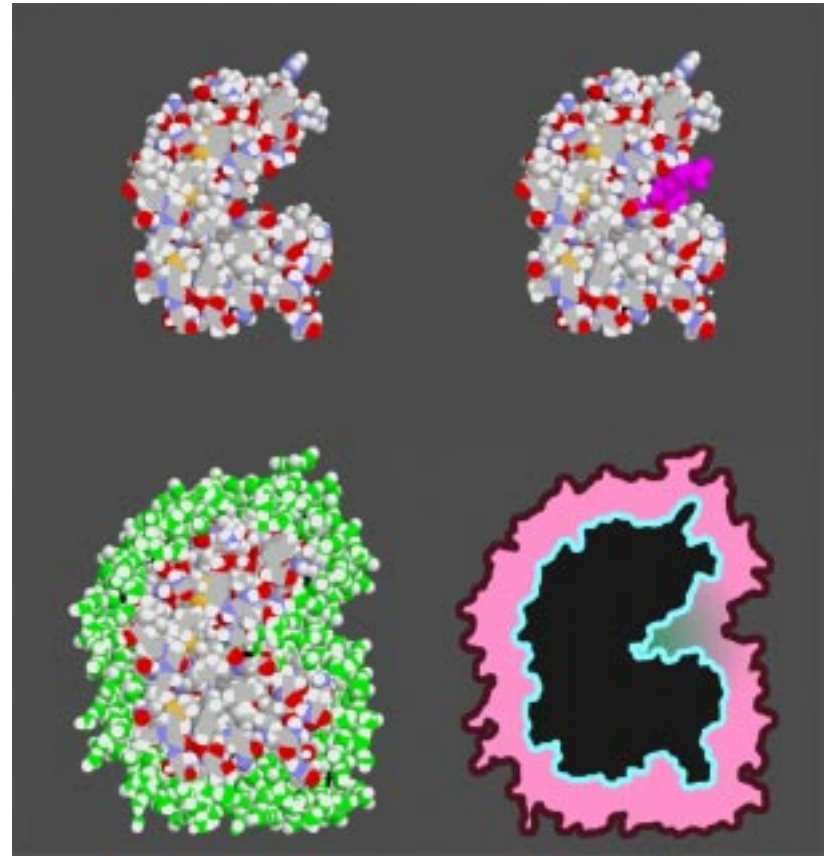
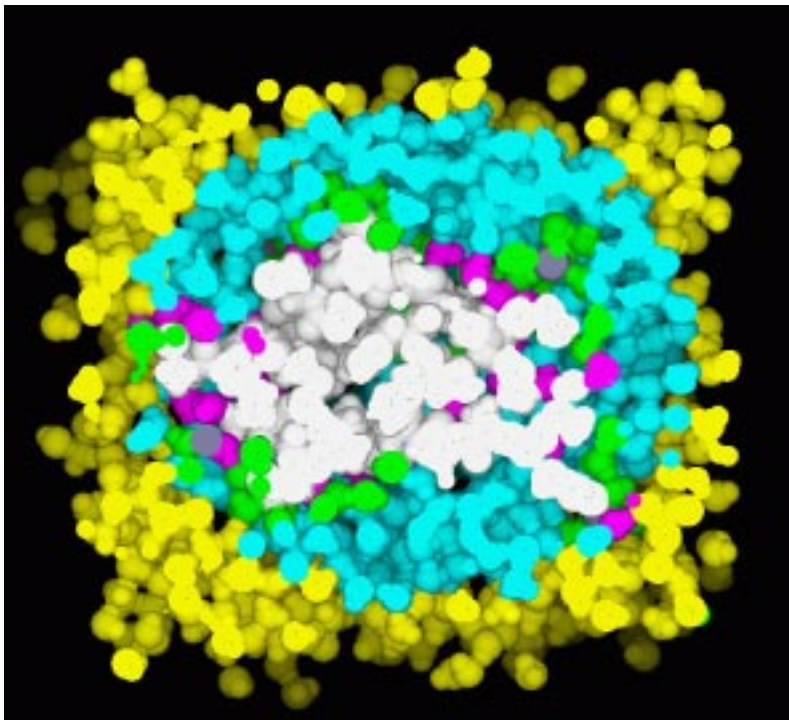
Water around Hydrophobic Groups on protein surface is more Compressible

- 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

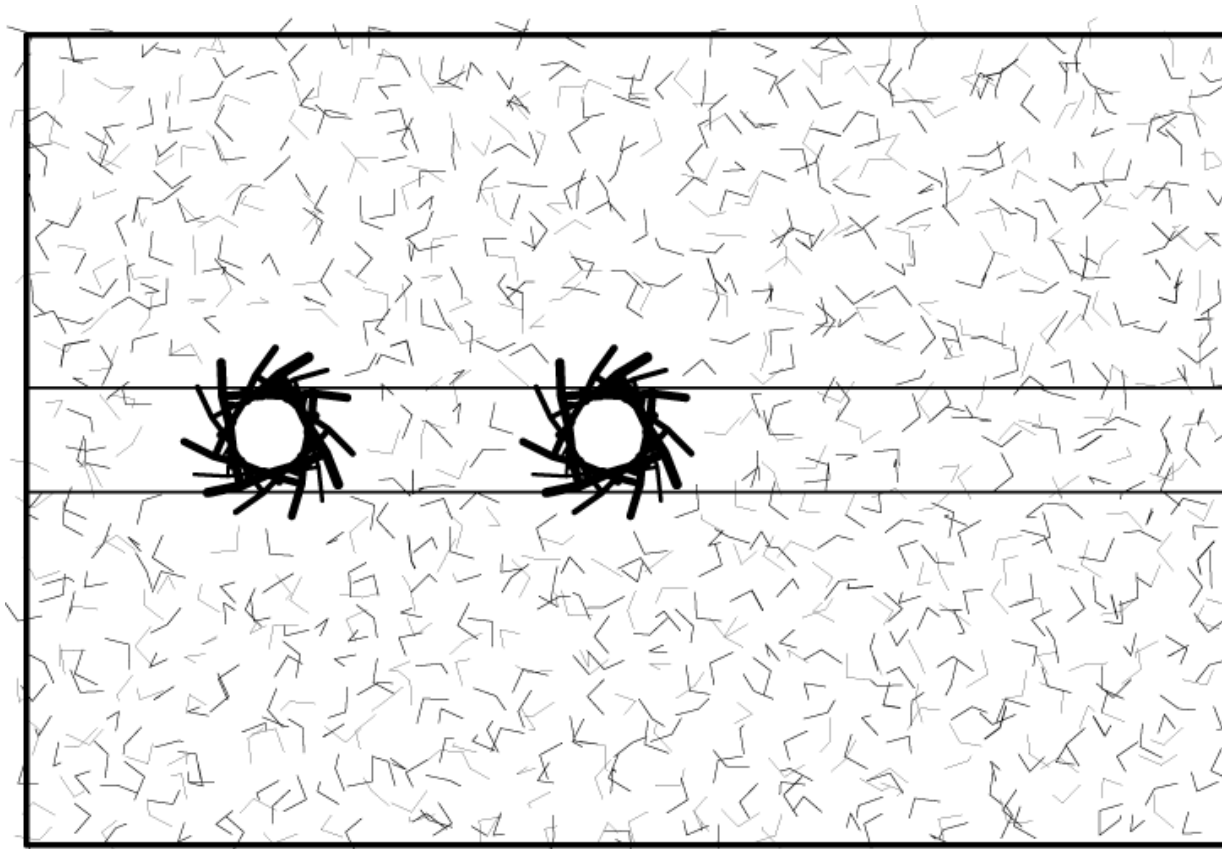
Interaction Between Water and the Protein Surface



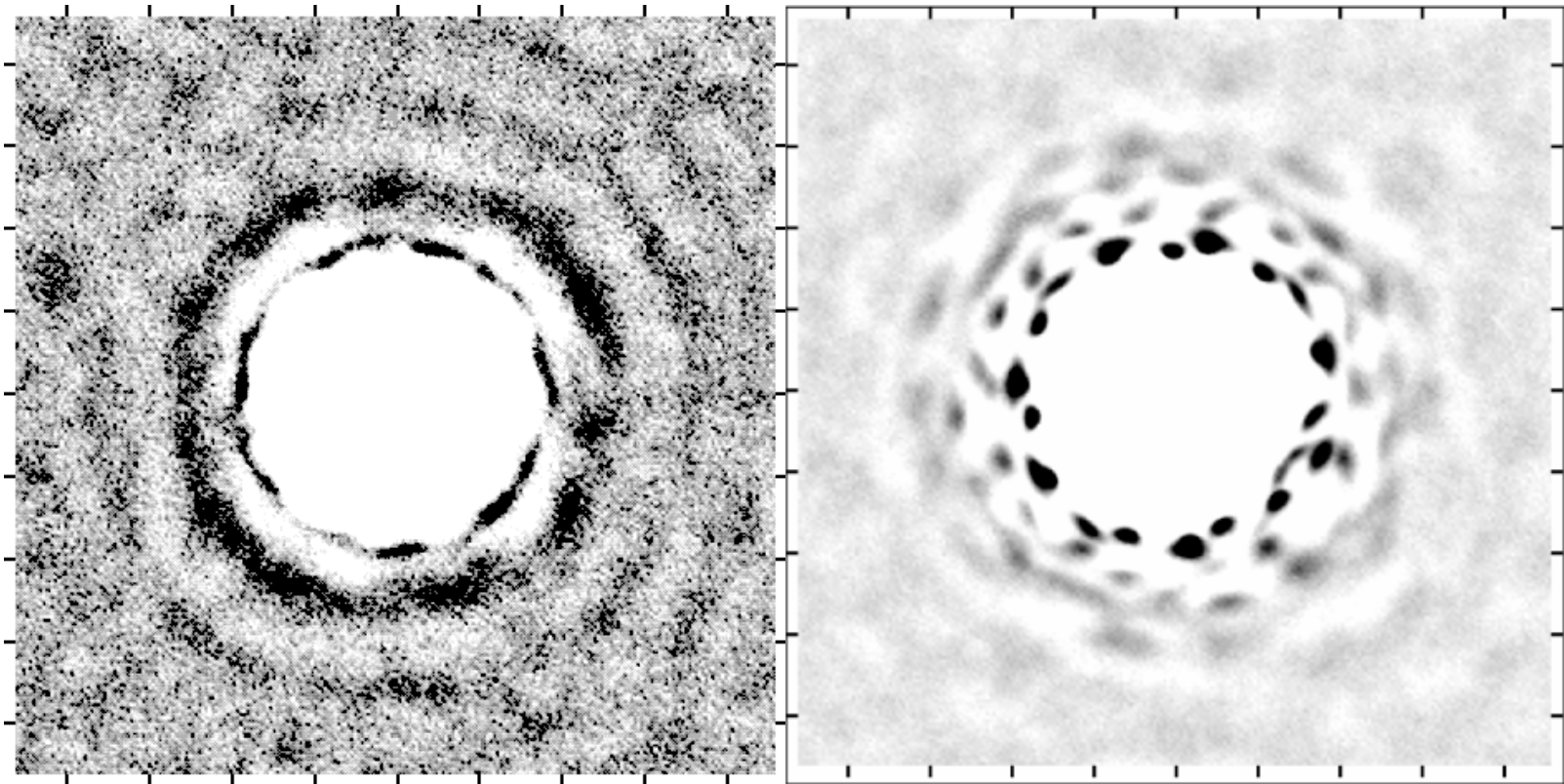
← THE PROTEIN SURFACE presents a very interesting interface from the point of view of water structure since it has a very irregular shape and has polar and non-polar atoms juxtaposed in close proximity. A slice through one frame of a simulation of water around a protein is shown. The protein is shown 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”)

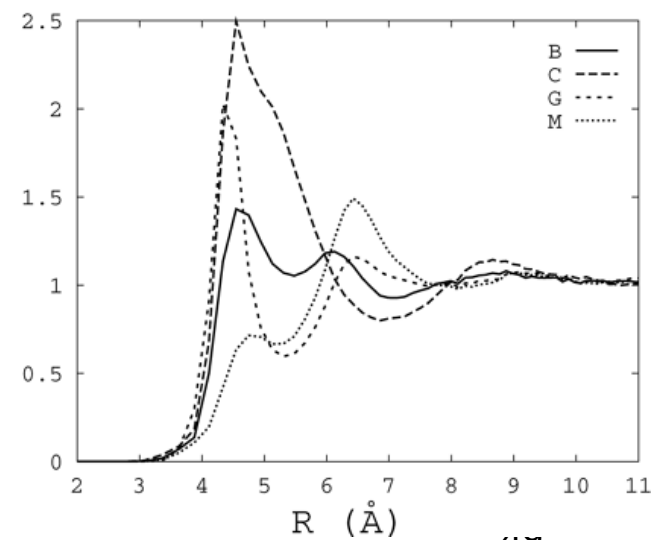
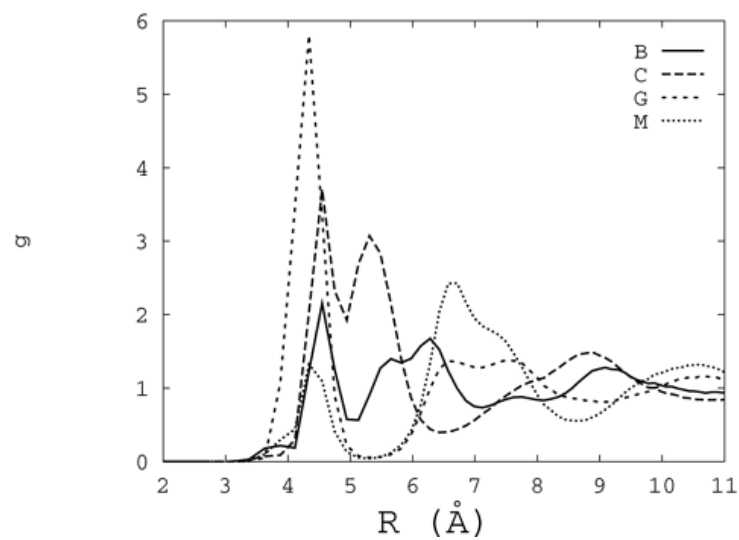
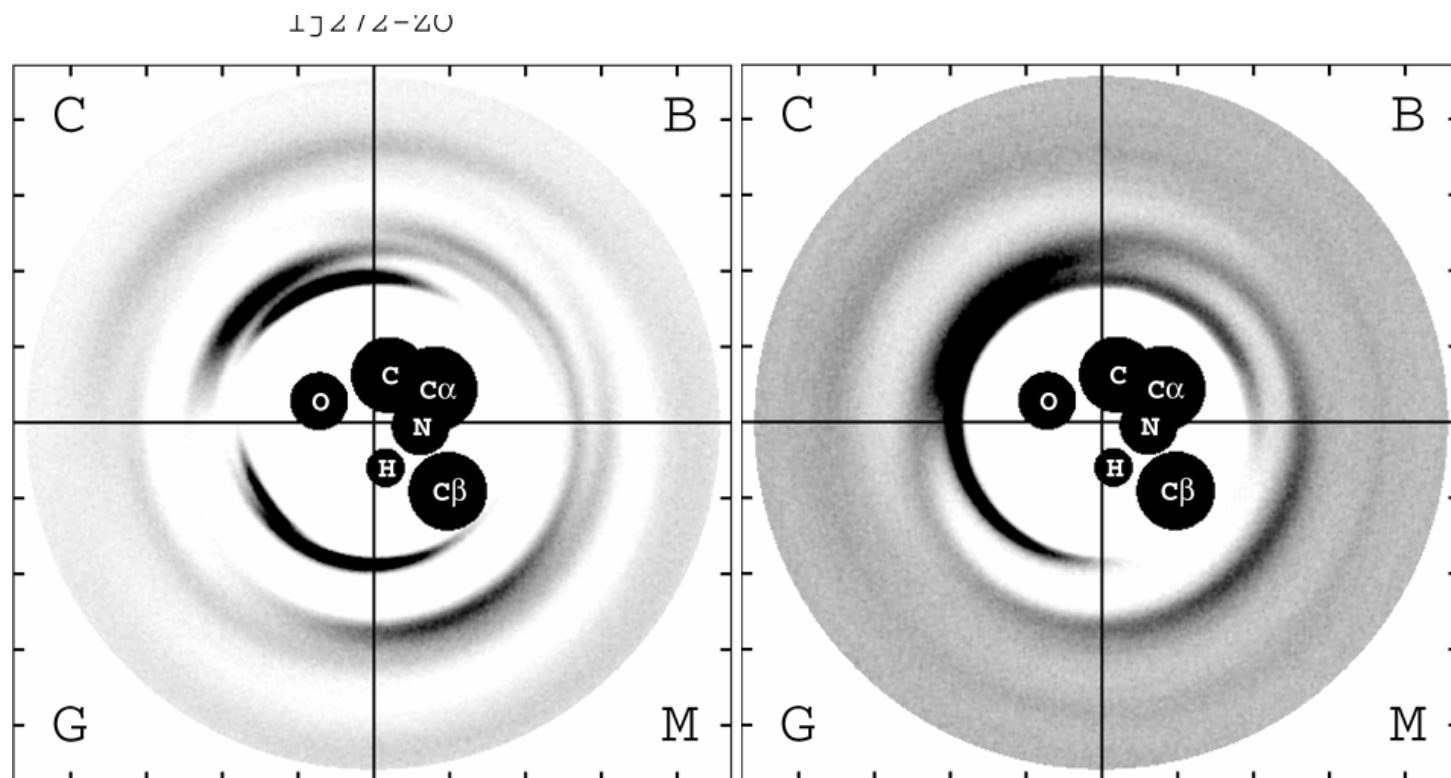


Second Solvent Shell: Water v LJ Liquid



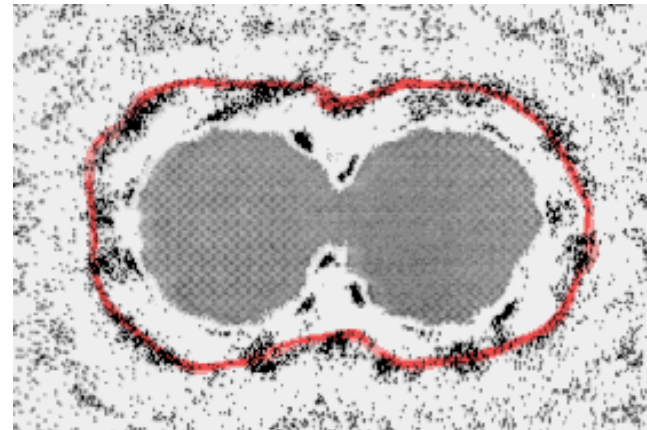
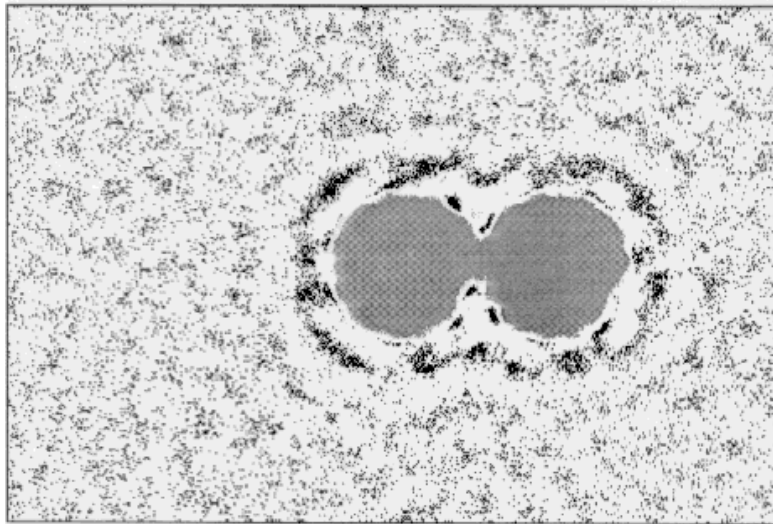
Water
vs.
Ar

(Helical
Projections)

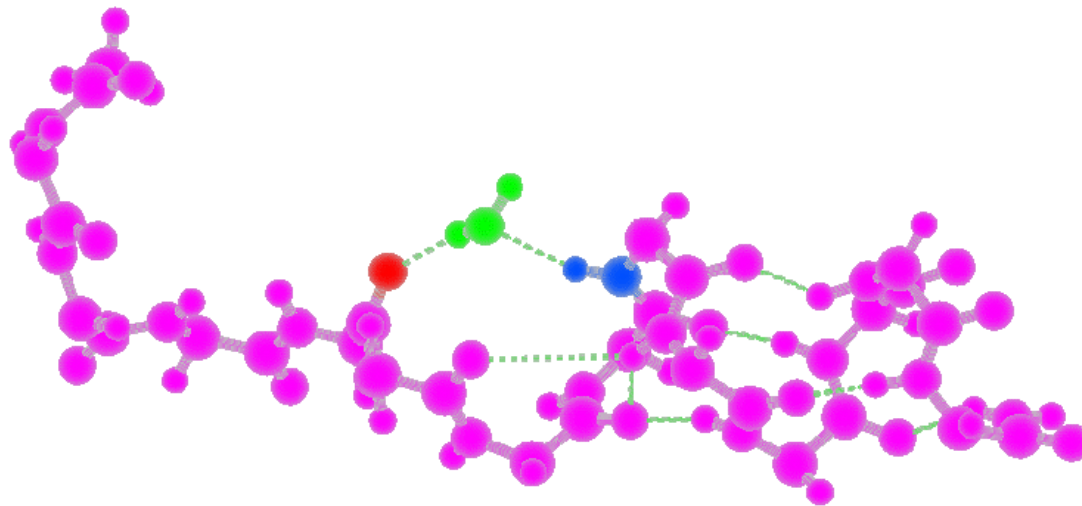


Hydration Surface

- 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.

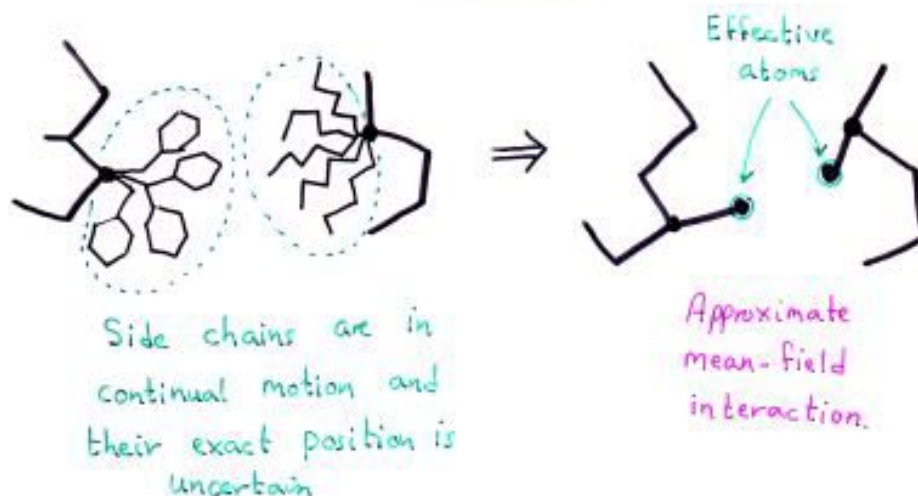


Illustration from M Levitt,
Stanford University

Simplified Protein: Lattice Models

- Cubic Lattice
- Tetrahedral Lattice

Illustration from M Levitt,
Stanford University

VERY SIMPLE LATTICE MODEL

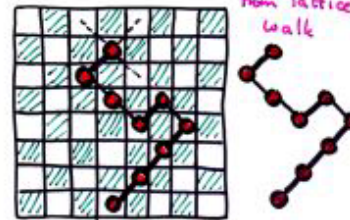
* Hinds & Levitt
J. Mol. Biol. 258, 201 (96)

DAVE HINDS

- Connect adjacent white squares on a chess board.

This gives:

- A chain
- Self-avoidance
- Bounds

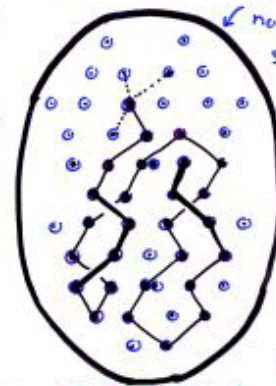


Get fold
from lattice
walk

- Connect adjacent vertices of a 50 vertex volume of a tetrahedral lattice.

This gives:

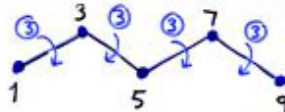
A bounded, relatively compact self-avoiding chain in 3-dimensions.



Ellipsoid but
no specific
shape

5 Å

At this resolution can represent real protein chain paths to 5 Å (coordinate Rms).



- Put every 2nd residue on the lattice $3^{n/2}$ folds.

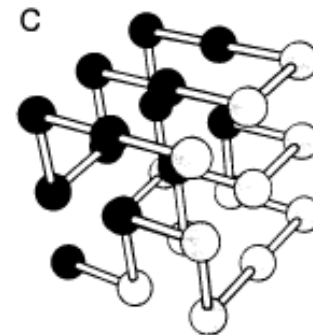
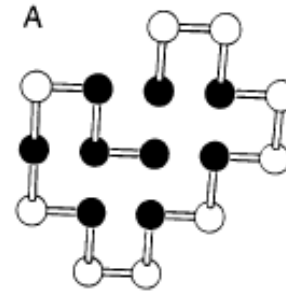
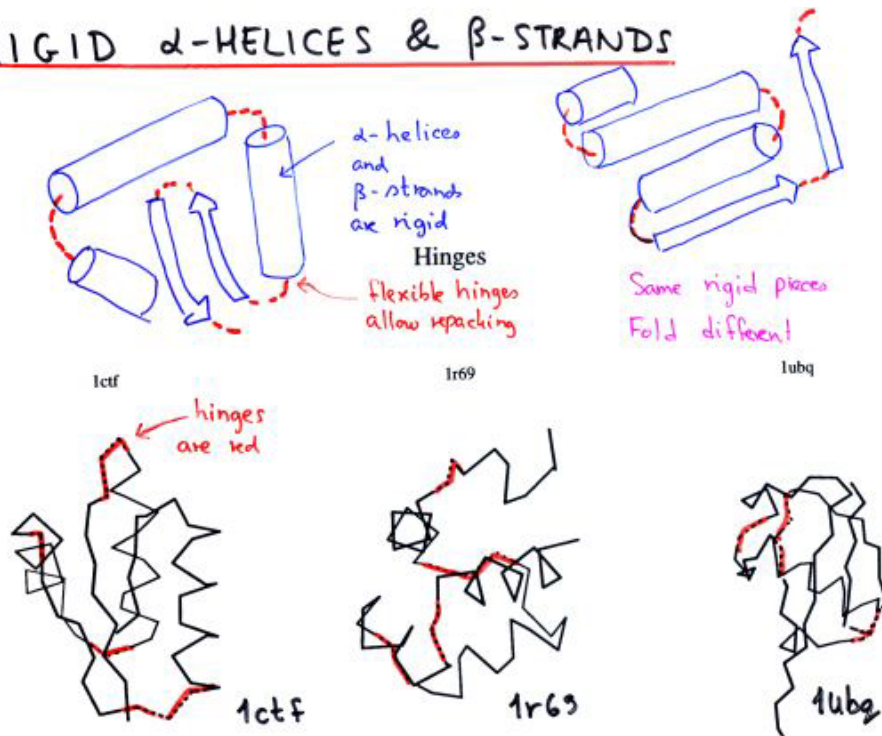


Illustration from
Dill et al. (1990)

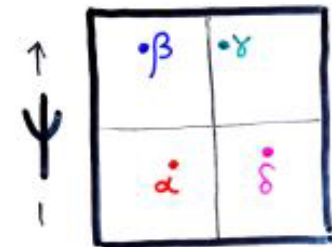
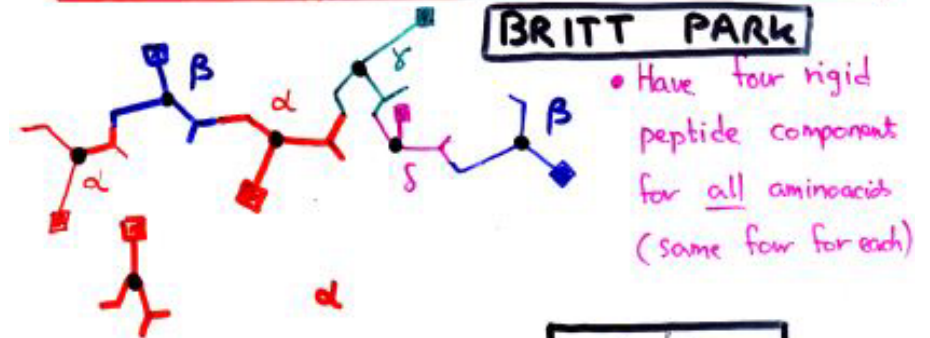
Off-lattice Discrete State Models

RIGID α -HELICES & β -STRANDS



FOUR-STATE OFF-LATTICE MODEL

BRITT PARK



- Like having $-\phi \rightarrow$ only four allowed (ϕ, ψ) 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?

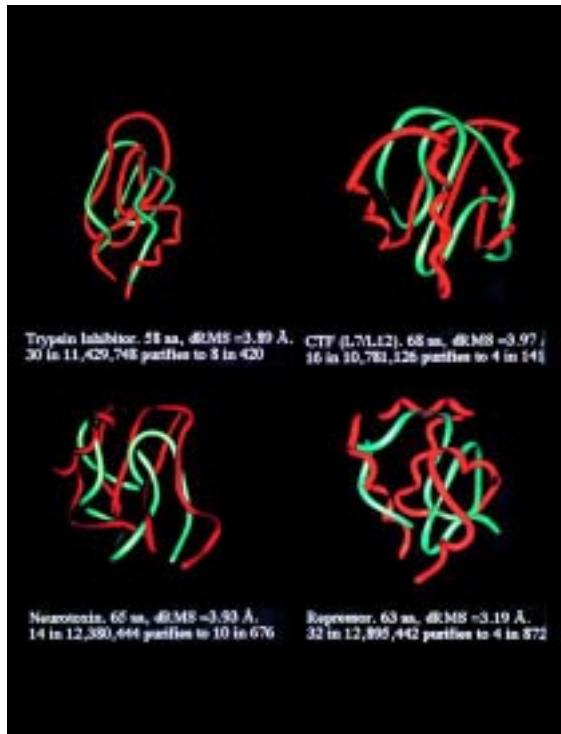


Illustration Credit: Dill et al. (1995)

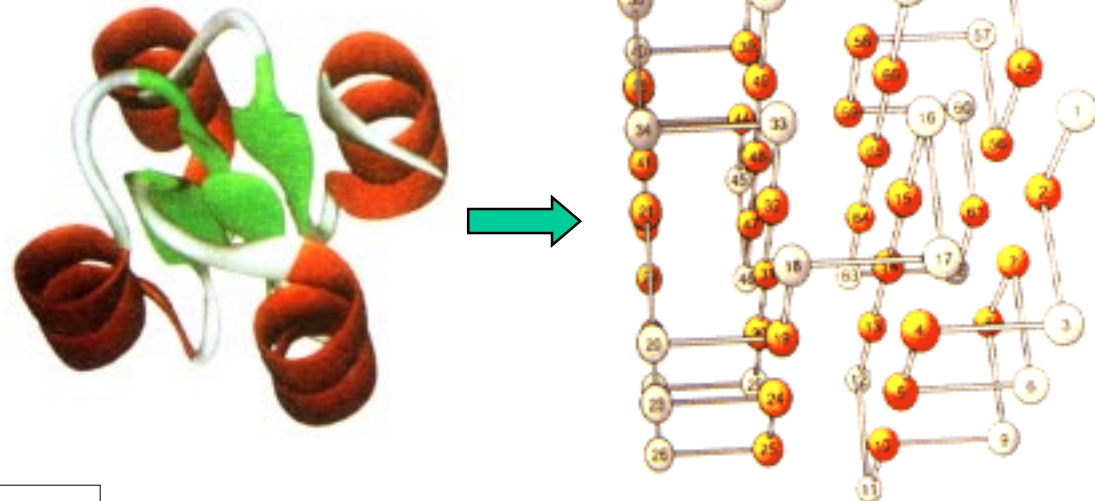
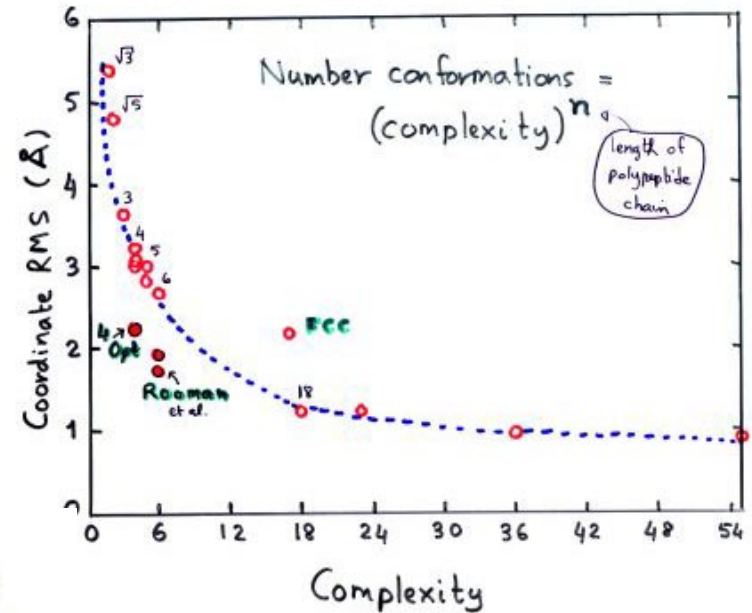


Illustration Credit: Hinds & Levitt (1992)

How well does the off-lattice model fit?



4-STATE MODEL FITS X-RAY WELL

Fits to X-Ray Structures

Best fit with only four states in dotted line

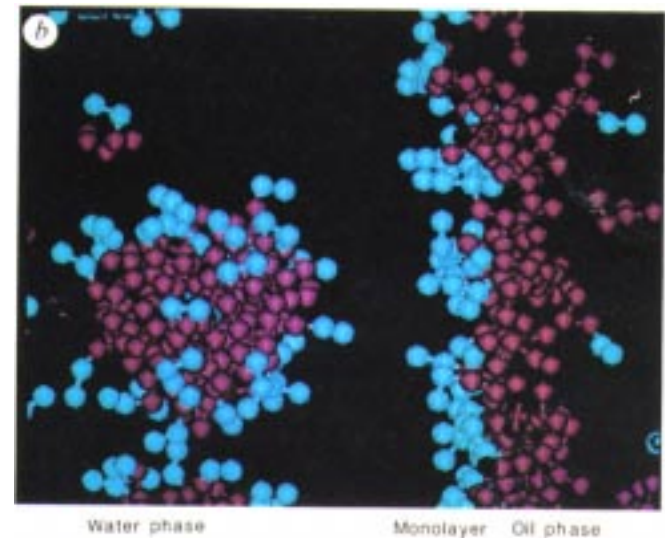
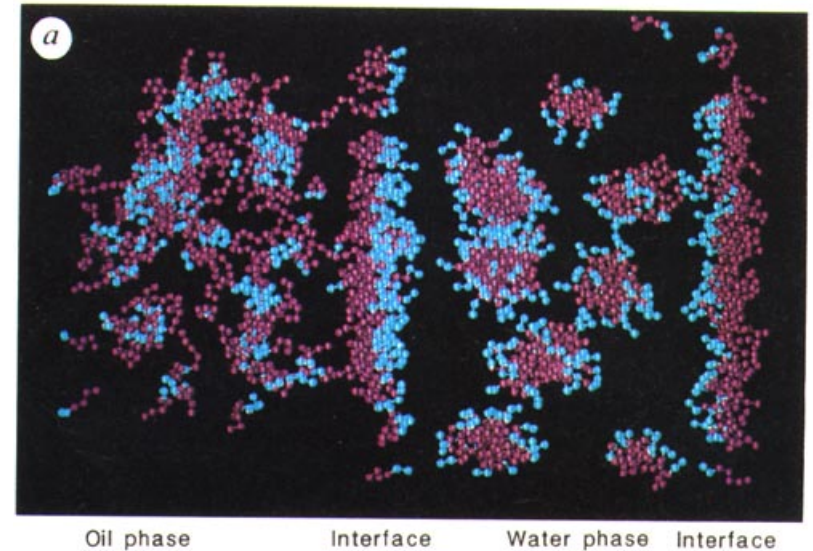


Model
Complexity vs
Fit to Reality

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.

	(S)imulation.	(B)asic Forces	(E)lectrostatics II
Preferred Lecture			
counts			

	(1) too simple	(2) just right	(3) too complex
Level of Simulation			
counts			

	(1) too simple	(2) just right	(3) too complex
Level of Basic Forces			
counts			

	(1) too simple	(2) just right	(3) too complex
Level of Electrostatics II			
counts			

Feedback on 2nd three computational lectures

- Which lecture did you like better
(‘S’ for Simulation,
‘B’ for Basic Forces,
‘E’ for Electrostatics II)?
- Was the simulation lecture at right level
(‘1’ for too basic, ‘2’ for just right, ‘3’ for too complex)?
- Was the basic forces lecture at right level
(‘1’ for too basic, ‘2’ for just right, ‘3’ for too complex)?
- Was the electrostatics (II) lecture at right level
(‘1’ for too basic, ‘2’ for just right, ‘3’ for too complex)?
- Sample responses: ‘S, 3, 2,1’ or ‘E-2-2-2’

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., Brucoleri, 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.