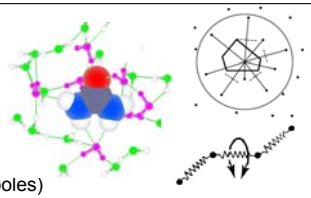
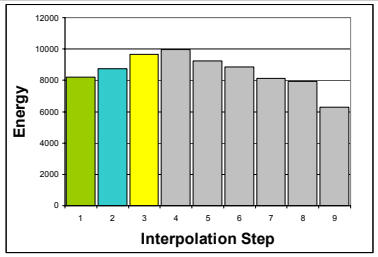
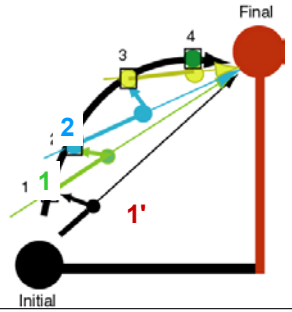
Forces, Simulation: the Outline



- The machinery of simulation
  - ◊ Electrostatics Review (polarization, multipoles)
  - ◊ Long-range Basic Forces are electrical: VDW Forces, electrostatic (H-bonds), Protein Packing in terms of VDW
  - ◊ Short range basic forces: QM bonds as springs
  - ◊ Basic Simulation Methods: minimization, MD, MC
  - ◊ Other Simulation Methods: Adiabatic Mapping, Normal Mode Analysis
- Doing simulation
  - Practical Aspects of Simulation: Programs & Potentials
  - Analysis: what can be Calculated from Simulation? (RDFs & Diffusion constants for liquids, RMS and E for proteins)
- Dealing with the Complexity of Proteins
  - Examples
  - Electrostatics Revisited: the Poisson-Boltzmann Equation
  - Water Simulation and Hydrophobicity
  - Simplified Simulation

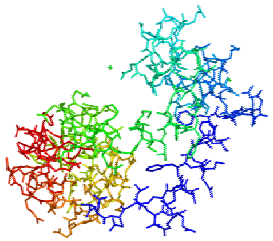
Pathway Interpolation via Adiabatic Mapping





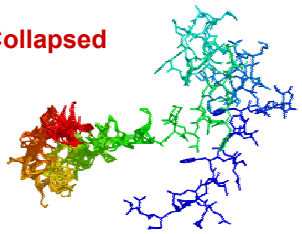
- **1'** Interpolation Step
- **1** Energy minimization (VDW + bonds)
- **2** Re-interpolate, re-minimize….
- * Slows down over humps

Adiabatic Mapping vs Linear Interpolation Strategies Compared with Calmodulin



Frame 4 (adiabatic)

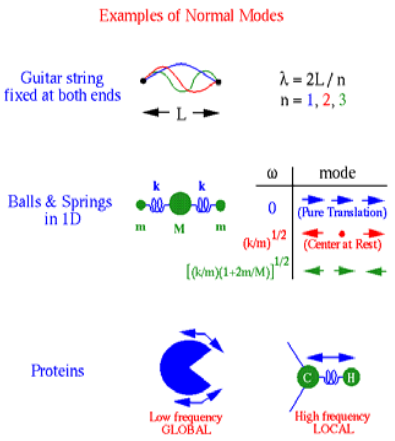
Collapsed



Frame 4 (linear)

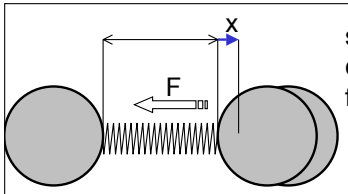
Normal Mode Analysis

- Normal Mode Analysis (NMA), also known as "Harmonic Analysis" is a classical technique for studying the vibrational and thermal properties of various molecular structures at the atomic level.
- It was originally developed for inorganic chemists wishing to identify peaks in the optical spectra (e.g., Fourier Transform Infrared (FTIR) and Raman spectra) of small molecules.
- Normal modes of vibration are simple harmonic oscillations about a local energy minimum, characteristic of a system's structure R and its energy function V(R).



Source: (adapted from W Krebs)

Normal-Mode Maths by Analogy to 1D springs



spring const k, freq. ω

$m\ddot{x} = -kx$

$(f = ma = -kx)$

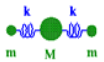
$x + Fx = 0$

$x(t) = A\sin(\omega t - \delta)$

PE $V = \frac{1}{2}kx^2$

$\frac{\partial^2 V}{\partial x^2} = k$

Generalization to many dimensional spring system (generalized displacement from equilibrium q_i)



$q_1 = \sqrt{m_1} \cdot (x_1 - x_{1e})$

$q_{3N} = \sqrt{m_N} \cdot (z_N - z_{Ne})$

$\ddot{\mathbf{q}} + \mathbf{F}\mathbf{q} = 0$

$\mathbf{F}\vec{\mathbf{A}} - \omega^2\vec{\mathbf{A}} = 0$

$q_i = A_i \cos(\omega t + \varphi)$

$\frac{\partial^2 V}{\partial q_i \partial q_j} = k_{ij}$

Harmonic Approximation in Anharmonic Systems and in real Proteins

- **For a theoretical systems with a purely harmonic V(R), there exists a superposition of normal modes that exactly expresses any given motion.** (A mathematician would say that the normal modes are basis vectors that span the set of all possible motions.)
- **For an anharmonic V(R) --- i.e. real world systems --- an harmonic potential still gives an excellent approximation near a potential energy minimum, and any small-amplitude motion around such a minimum can still be well described by a sum of normal modes.**
- Any classical system can be said behave harmonically at a sufficiently low temperature.
- A typical normal mode analysis computes the characteristic vibrations and the corresponding frequencies assuming V(R) is harmonic in all degrees of freedom.
- Only systems near a potential energy minimum exhibit harmonic behavior, so normally, the system is first energy-minimized system (T=0+ K) to assist with the harmonic assumption.
- Around 200 K, the fluctuations in the atoms of a globular protein begin to deviate considerably from harmonic behavior. As the protein is heated to 300 K, its motion becomes considerably anharmonic.
- These facts should be kept in mind when attempting to interpret *in vivo* and *in vitro* protein behavior using normal modes.

Computational Costs of NMA

- **Normal mode analysis is less expensive than Molecular Dynamics (MD) simulation, but because the computer must invert large matrices**, requires much more memory when dealing with large molecules.
- This problem can be overcome somewhat by clumping regions, such as amino acid residues, and treating them as if they were a single atom, effectively reducing the number of atoms, and hence the size of the matrices the computer must invert.

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)

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

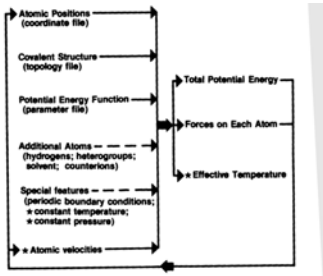


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of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

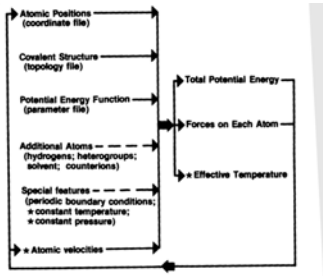
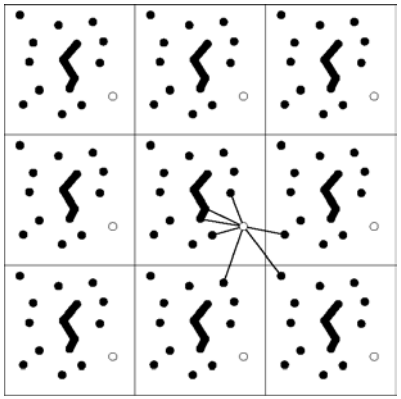


Fig. 4.1. Schematic flow chart of algorithms for energy minimization and molecular dynamics. Features which apply only to molecular dynamics are indicated by asterisks. Dashed lines indicate optional input. Each cycle of energy minimization represents a step in conformation space, while each cycle of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

Illustration Credit: McCammon & Harvey (1987)

Periodic Boundary Conditions

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

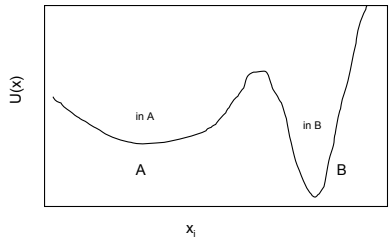


Average over simulation

- Deceptive Instantaneous Snapshots (almost anything can happen)
- Simple thermodynamic averages
  - ◊ Average potential energy <U>
  - ◊ $T \sim <\text{Kinetic Energy}> = \frac{1}{2} m <v^2>$
- 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 = Sum over all P_i , to normalize probabilities to 1



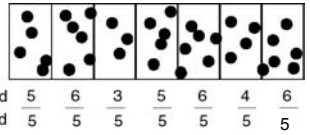
- Entropy
  - ◊ $S(A) = k \sum (P_i \ln P_i)$, where the sum is over points i in A
- Free Energy
  - ◊ G(A) = U(A) - TS(A)
- Entropy and Free Energy are only defined for distinctly diff. "states" -- e.g. A ("unfolded")and B ("folded")
  - ◊ State B has a lower U and its minimum is more probable than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

Motion	length (Å)	time (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

Number Density

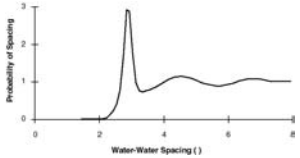


= Number of atoms per unit volume averaged over simulation divided by the number you expect to have in the same volume of an ideal "gas"

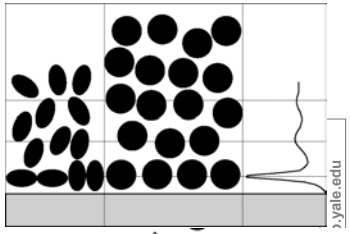
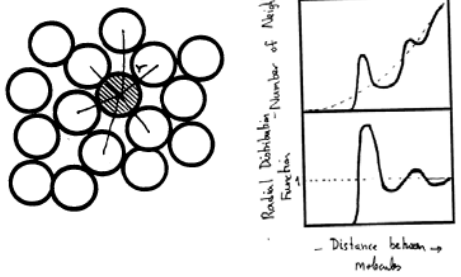
Spatially average over all directions gives

1D RDF =

[Avg. Num. Neighbors at r]
[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