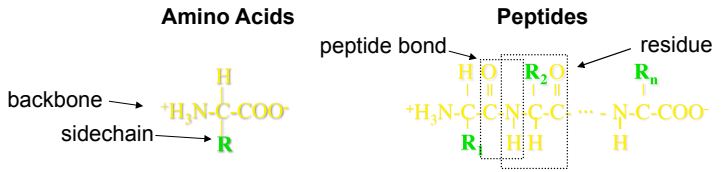
BIOCH 590: Biomacromolecules Spring 2010

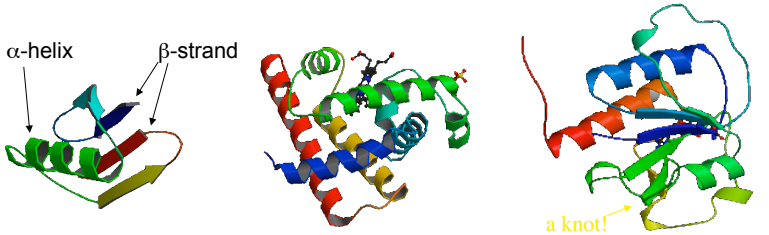
Molecular Interactions and Biomolecular Structures; Computer Modeling

Main References: 1. Chapter 3 of van Holde
2. Second half of Chapter 9 of Tinoco (pages 493-516)

Hierarchical Organization of Proteins



Primary Sequence: MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE

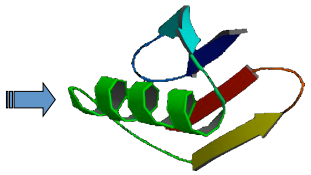


Protein G B1 (3gb1) **Myoglobin (1mbc)** **tRNA transferase (1mxi)**

Structure-Function Paradigm

Sequence Folded 3D Structure Functions

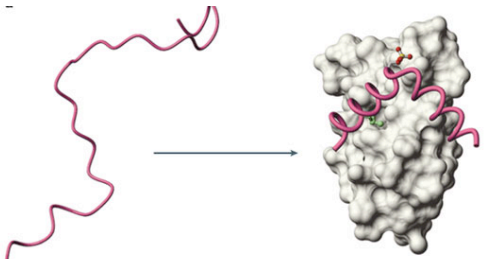
MTYKLILNGK
TLKGETTTEA
VDAATAEKVF
KQYANDNGV
DGEWTYDDA
TKTFTVTE …



catalysis
transport
signaling
motors …

- Structural Genomics (NMR, X-ray)
- Protein structure predictions

Intrinsic Disorder & Cellular Functions



Dyson & Wright, *Nat Rev Mol Cell Biol* (2005)

- Intrinsically Disordered Proteins (IDPs): functional proteins that can exist as dynamic ensembles of disordered structures *under physiological conditions*.

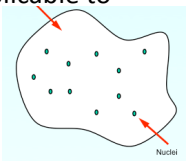
- Over 30% of eukaryotic proteins are predicted to be disordered.
- Critical in cellular regulation and signal transduction.
- Allow high specificity with low affinity; structural plasticity for binding diversity
- Particular importance in cancer and protein misfolding diseases.

Determinant of Structure (or Lack of It)

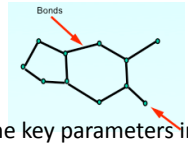
- Probability of observing a particular structure (conformation) is determined by its stability (as defined by the free energy)
  - Thermodynamics and statistical mechanics!
- No single structure is *the* structure
  - It is all about probability (statistical mechanics!)
  - Motions and flexibility are important too
- The stability depends on a range of factors
  - Intramolecular interactions
    - Bonded: chemical bonds, angles, dihedrals etc
    - Nonbonded: "weak" interactions
      - Charged-charged, van der Waals (dispersion and repulsion)
  - Intermolecular interactions: nonbonded/weak interactions
    - Cellular environment: solvent (water), membrane, salt, pH etc
    - Association with other biomolecules, small molecules, ions, etc

Quantum Mechanics vs. Molecular Mechanics

- Quantum mechanics: "exact" and most applicable to understand chemical reactions
  - Separate nuclei and electrons
  - Too expensive, and not sufficiently accurate
  - Not relevant as many biological processes



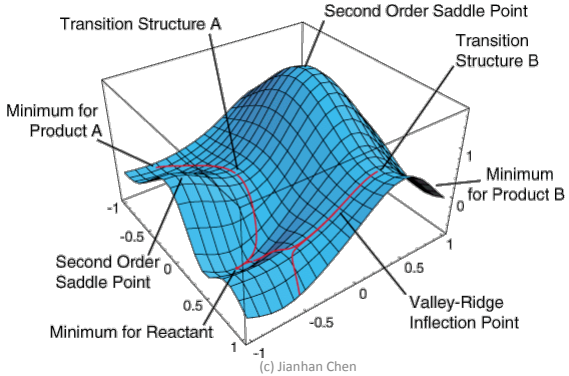
- Molecular mechanics: classical mechanics at molecular level
  - Classical treatment of all atoms
  - No electron, no chemistry
  - Allows description of large molecules
  - Experimental methods available to determine the key parameters in a molecular mechanical treatment



- Hybrid QM/MM
  - QM for the active site (where reaction occurs) and MM for the rest
  - Accurate treatment of MM/QM Boundary is a problem

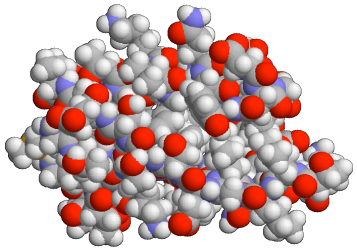
Classical Mechanics

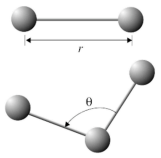
- Total energy: $E = K + V$
  - Kinetic energy ($K = mv^2/2$), potential energy V (i.e., force field)
- Newton's second law of motion: $F = m\,a$
  - Relation of force and potential energy: $F = -\delta V/\delta r$

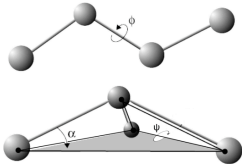


Molecular Potentials

- Basic form: $V = V_{bonding} + V_{nonbonding}$
  $= (\Sigma V_{bond} + \Sigma V_{angle} + \Sigma V_{dihe}) + \Sigma(V_{elec} + V_{vdw})$
  - The potential energy is a function of all coordinates.
  - Additivity, empirical, transferability

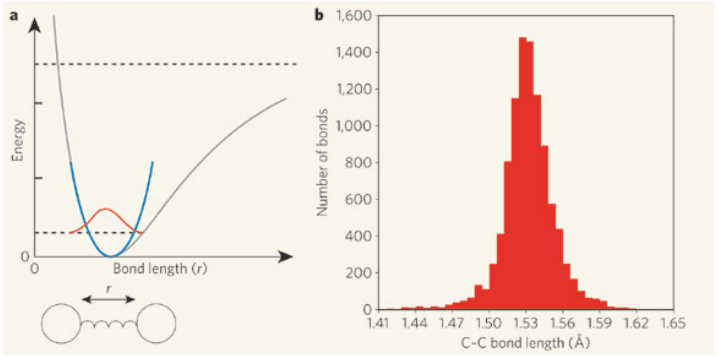






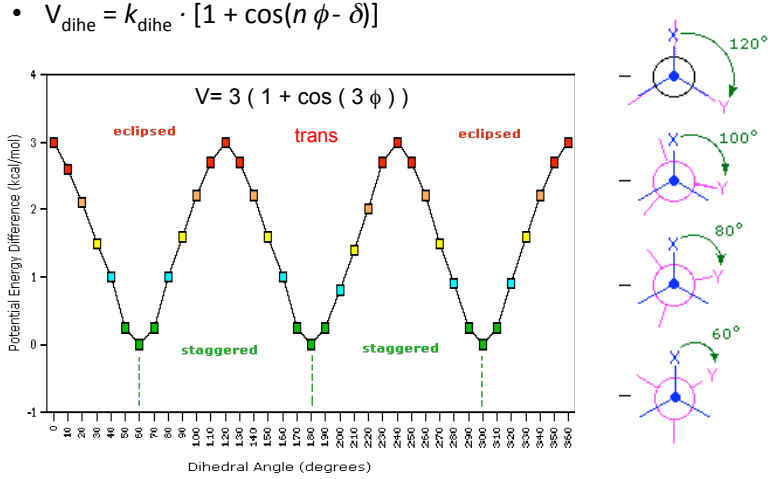
Bonds and Angles

- $V_{bond} = k_{bond} (r - r_o)^2$
  - Harmonic approximation
  - OK for biomolecules
- $V_{angle} = k_{angle} (\theta - \theta_o)^2$



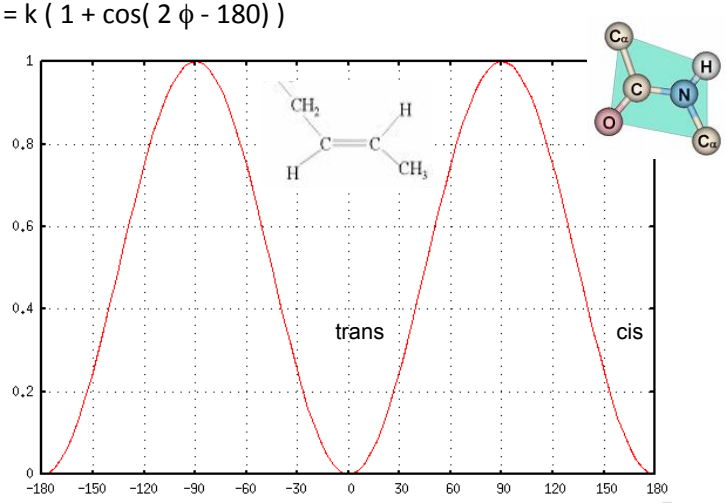
Dihedral Potentials

- $V_{dihe} = k_{dihe} \cdot [1 + \cos(n\phi - \delta)]$



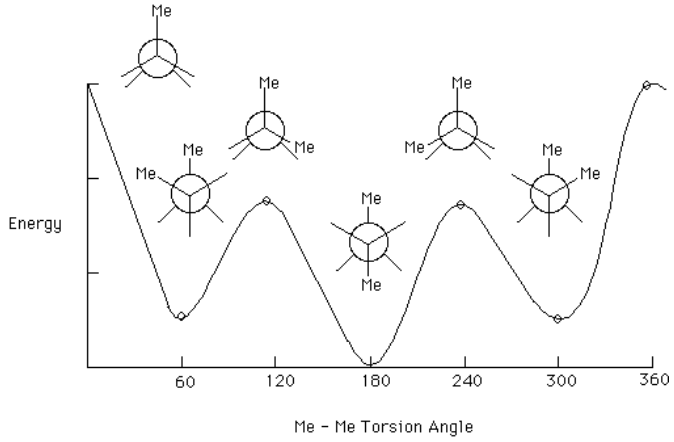
Double Bonds: *n*=2, *δ*=-180

- V = k (1 + cos(2 ϕ - 180))



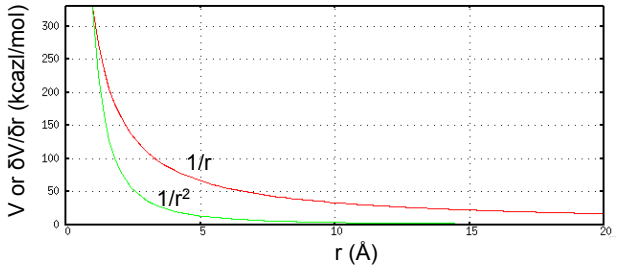
Realistic Dihedral Potentials

- Actual dihedral potentials often have contributions with multiple periodicities



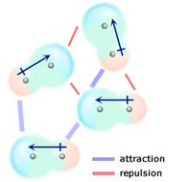
Electrostatic Interactions

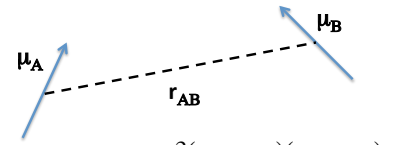
- $V_{elec} = q_1q_2/4\pi\varepsilon_0 r$ Coulomb's Law
  - ε_0: permittivity constant of vacuum
- A simplified form: $V_{elec} = 332\ q_1q_2/r$
  - Where *q* is unit of electron charge, *r* is in Å and *V* in kcal/mol.
- Dielectric medium: $V_{elec} = 332\ q_1q_2/\varepsilon r$
  - ε is dielectric constant (relative permittivity).
  - ε=78 for water under lab conditions (300K, 1atm)



Dipole-Dipole Interactions

- Dipole moment: arise from charge separations
  - measure the "polarity" of a molecule (or fragment)
- Dipole-dipole interactions
  - Fully included if all charges treated explicitly
  - Offer simplifications (by reducing the number of terms)



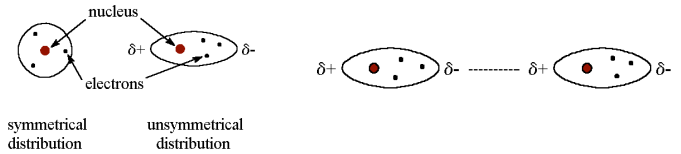


$$V_{dd} = \frac{\mu_A \bullet \mu_B}{|r_{AB}|^3} - \frac{3(\mu_A \bullet r_{AB})(\mu_B \bullet r_{AB})}{|r_{AB}|^5}$$

  - Decays faster: $1/r^3$ dependence
  - Special cases: parallel and perpendicular orientations

van der Waals Interactions

- London dispersion: attractive forces that arise from temporary dipoles (induced dipole-induced dipole interactions)

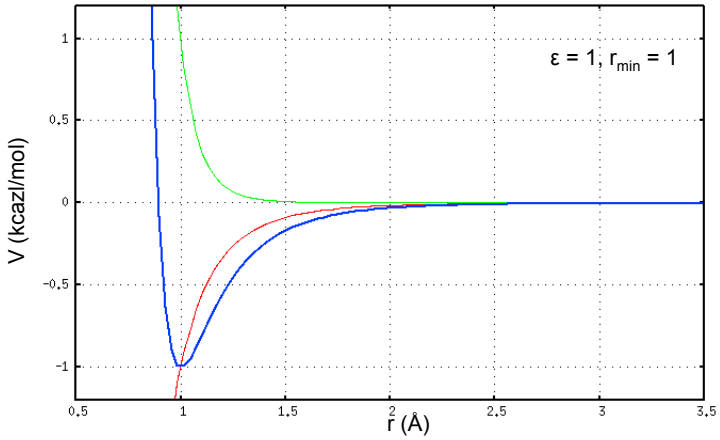


- van der Waals repulsion: all atoms repel at short distances
- A common function form: $V_{vdw} = -A/r^6 + B/r^{12}$
- Lennard-Jones potential function (12-6)

$$V(r) = \epsilon\left[\left(\frac{r_{min}}{r}\right)^{12} - 2\left(\frac{r_{min}}{r}\right)^{6}\right]$$

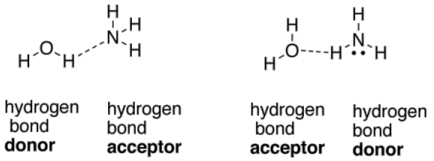
Lennard-Jones Potential

$$V(r) = \epsilon\left[\left(\frac{r_{min}}{r}\right)^{12} - 2\left(\frac{r_{min}}{r}\right)^{6}\right]$$



Hydrogen Bonds

- Very important in macromolecule structures
- Primarily a dipole-dipole interaction, but arguably with some covalent nature (electron sharing in so-called low barrier HBs)

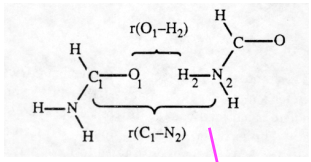


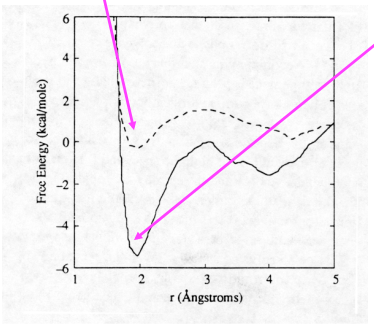
- The strength of HBs vary greatly and depend on the environments (dielectric screening)
- The functional form for HB is unclear.
  - Often mimicked by Lennard-Jones potential
  - At present, often treated with electrostatic + vdW

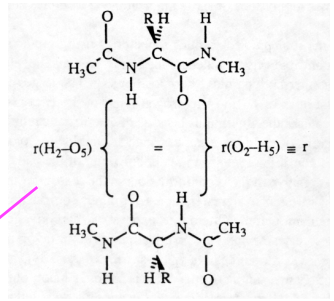
What is a hydrogen bond worth?

Secondary Structure	Stability per H-bond	Model	Reference State
Antiparallel β-sheet	-2.8	$[Ac\text{-}ala\text{-}NHMe]_2$	Infinite separation
Ala-gly Type II turn	-0.6	Ac-ala-gly-NHMe	Extended
Amide H-bond	-0.3	$[formamide]_2$	Infinite separation
1st helical H-bond	-0.2	$Ac\text{-}(ala)_3\text{-}NHMe$	extended
2nd helical H-bond	-0.4/-1.0	$Ac\text{-}(ala)_4\text{-}NHMe$	extended
Ala-gly Type I turn	2.6	Ac-ala-gly-NHMe	extended
Pro-gly Type I turn	2.6	Ac-pro-gly-NHMe	extended

β-sheets can have exceptional stability





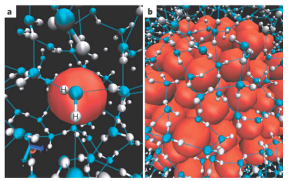


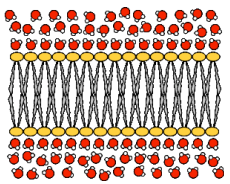
Small barrier and minimum associated with "naked" hydrogen bond, much more significant for beta-sheet model

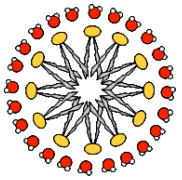
Tobias and Brooks CPL (1990)

Hydrophobic Effects

- The property that nonpolar solutes aggregate in water
- Arise from a combination of elemental physical effects
  - Difference in strengths solute-water and water-water interactions
  - Difference in shapes (sizes) of solutes and water
  - Various entropic contributions
- One of the key driving forces for self-assembly in biology
  - Biological membrane, micelle formation, protein folding …
  - Complex temperature dependence: cold denaturation of proteins
  - Very difficult to describe theoretically!



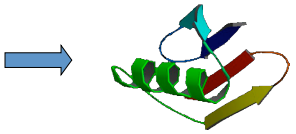




The Protein Folding Problem

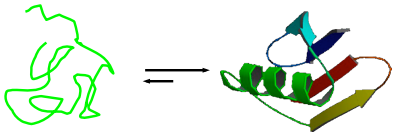
- How does the primary Sequence specify the native fold?
  (Afinsin, Science, 1973)

MTYKLILNGK TLKGETTTEA
VDAATAEKVF KQYANDNGVD
GEWTYDDATK TFTVTE



- Levinthal's Paradox: how protein folds this fast?

τ_{fold} < 1 second



10 conformations per residue,
10^{-11} (10 ps) per state

10^{60} states for a 60-residue protein,
$>10^{40}$ years for random search !

Determinant of Protein Structure

- Principles known, but a long way to go in quantification
- Key stabilizing physical interactions (hydrogen bonds, hydrophobic effects, etc) known
  - Relative "importance" and quantification not clear
- Reasonable success in practice of structure prediction and computational protein design
  - Largely based on statistical knowledge of known structures
- Much more to learn
  - Quantitative nature
  - Mechanisms of folding
  - Misfolding and aggregation
  - Protein dynamics

http://predictioncenter.org/

Biological Energy Scales

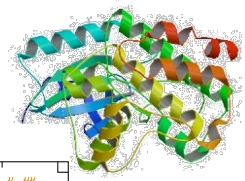
- Chemical bonds: C-H -105 kcal/mol, C=C 172 kcal/mol
- Ionic hydration: Na^+ -93 kcal/mol, Ca^{2+} -373 kcal/mol
- Hydrogen bonds: O…H -5 kcal/mol (*in vacuum*), can be much weaker in solution or protein environment
- Protein stability: ~ -2-10 kcal/mol (*in solution*)
- Protein-DNA binding: ~ -5-20 kcal/mol (*~200 Å² contact*)

A quantitative model of proteins need to be able to predict several kcal/mol differences in (free) energy!

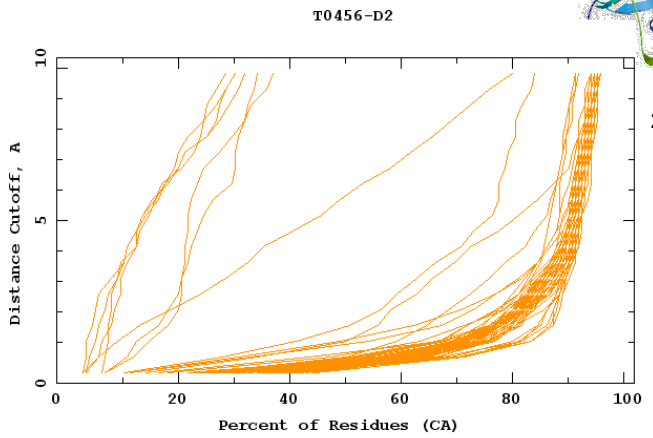
This small energies result from a summation of many atom pairs (10^6-10^8)! Plus, entropic effects.

This is extremely difficult, if ever possible!

CASP8 Target T0456

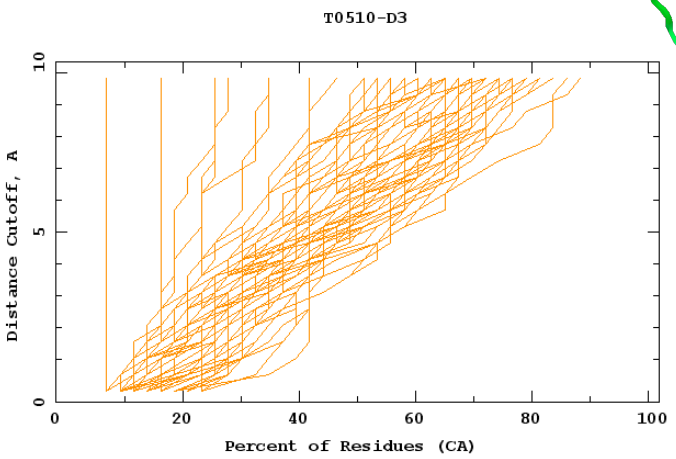


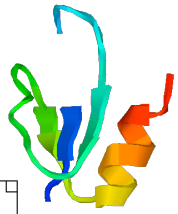
PDB: 3dfa
286 residues



http://www.predictioncenter.org/casp8/results.cgi

CASP8 Target T0510



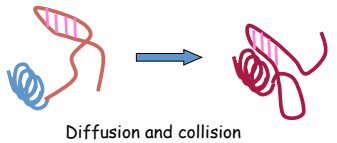


PDB: 3doa
44 residues

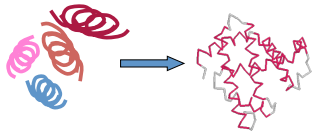
http://www.predictioncenter.org/casp8/results.cgi

How proteins fold?

- "Classical" understandings
  - Diffusion-collision
  - Hierarchical folding
  - Assembly of foldons

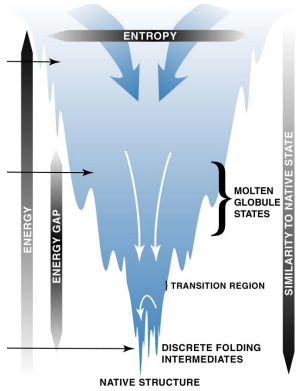


Diffusion and collision
Karplus and Weaver, *Biopolymers.*, 18, 1421 ('77).



Folding via modular assembly
Ptitsyn and Rashin, *Biophys. Chem.*, 3, 1 ('75).

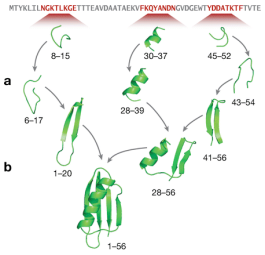
The Energy Landscape Theory
(arguably the prevailing theory)

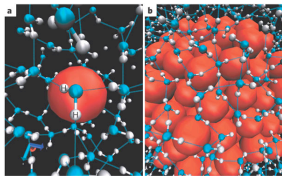


Wolynes et al., *Science* (1995).

Two Excellent Readings

- "The protein folding problem", Dill et al., Annu. Rev. Biophys. 2008, 37:289-316
  - A good overview of current understanding of protein folding
  - The zipping and assembly hypothesis is interesting, but as a prediction method the success has been limited
- "Interfaces and the driving force of hydrophobic assembly", Chandler, Nature 2005, 437:640-647
  - One of the most readable and informative reviews on current understanding of hydrophobic effects



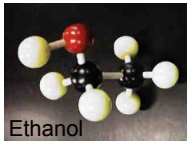


Basic Components of Modeling

Force Field

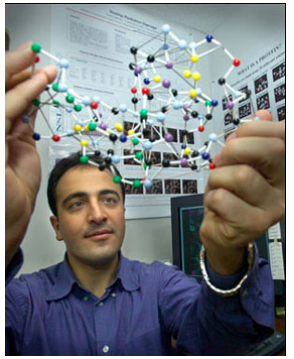
A set of basic model units and associated rules.

A pioneering ball and stick atomic model set, 1860s



Sampling

The process of finding the optimal assembly of the basic model units.

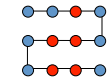


http://www.sesame.org.jo/publication/NSLS.aspx

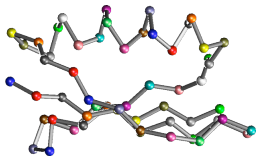
Molecular Models at Multiple Scales

increasing details and predicting power
increasing difficulty and computational cost

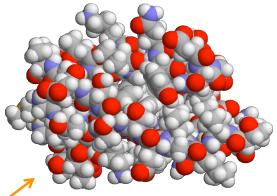
Lattice models	minimalist models	coarse-grained models	all-atom / simplified potentials	all-atom / molecular mechanics	quantum mechanics



polar
hydrophobic



Minimalist models



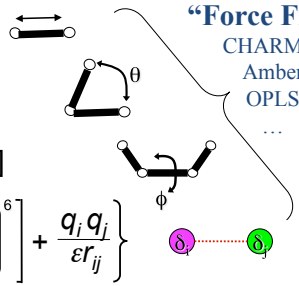
All-atom models

Provide the ultimate details necessary for understanding most biophysical processes.

Molecular Mechanics

Classical Energy Functions

$$V_{MM} = \sum_{bonds,i} \frac{1}{2} k_i^b \cdot (b_i - b_i^0)^2$$
$$+ \sum_{angles,i} \frac{1}{2} k_i^\theta \cdot (\theta_i - \theta_i^0)^2$$
$$+ \sum_{torsions,i} k_i^\phi \cdot [1 + \cos(n_i \phi_i - \delta_i)]$$
$$+ \sum_{atoms,i} \sum_{i<j} \left\{ \varepsilon_{min}^{ij} \left[\left(\frac{r_{min}^{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{min}^{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\}$$



"Force Field"
CHARMM
Amber
OPLS
…

Molecular Dynamics (MD)

$$m_i \ddot{r}_i = F_i = -\nabla V_i$$

Monte Carlo (MC)

$$P(\delta r) = \exp(-\Delta V / kT)$$

CHARMM param22 Force Field

- Topology file: define the building blocks (atoms, connectivities)

atom types

```
MASS     1 H      1.00800 ! polar H
MASS     2 HC     1.00800 ! N-ter H
MASS     3 HA     1.00800 ! nonpolar H
```

name type charge

residue blocks

```
…
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |     HB1
ATOM HA   HB      0.09  !     |    /
GROUP                   !  HA-CA--CB-HB2
ATOM CB   CT3    -0.27  !     |    \
ATOM HB1  HA      0.09  !     |     HB3
ATOM HB2  HA      0.09  !   O=C
ATOM HB3  HA      0.09  !     |
GROUP                   !
ATOM C    C       0.51
ATOM O    O      -0.51
```

atom compositions

connectivity

```
BOND CB CA  N  HN  N  CA  O  C
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
IMPR N -C CA HN  C CA +N O
DONOR HN N
…
```

excerpted from: top_all22_prot.inp

CHARMM param22 Force Field

- Parameter file: define the parameters of interactions

```
…
BONDS
C    C     600.000     1.3350 ! ALLOW ARO HEM
                ! Heme vinyl substituent (KK, from propene (JCS))
CA   CA    305.000     1.3750 ! ALLOW   ARO
                ! benzene, JES 8/25/89
…
ANGLES
CA   CA   CA    40.000    120.00   35.00   2.41620 ! ALLOW   ARO
                ! JES 8/25/89
CE1  CE1  CT3    48.00    123.50   !
                ! for 2-butene, yin/adm jr., 12/95
…
DIHEDRALS
C    CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP
                ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    CT2  NH1  C        0.2000  1   180.00 ! ALLOW PEP
                ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
…
NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 13.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
                !adm jr., 5/08/91, suggested cutoff scheme
C      0.000000  -0.110000     2.000000 ! ALLOW   PEP POL ARO
                ! NMA pure solvent, adm jr., 3/3/93
CA     0.000000  -0.070000     1.992400 ! ALLOW   ARO
                ! benzene (JES)
```

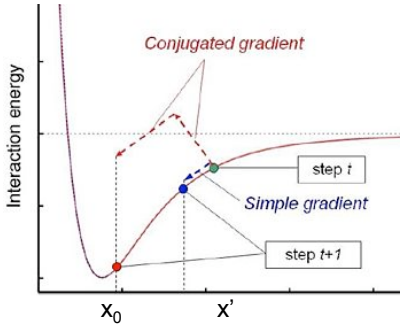
excerpted from: par_all22_prot.inp

Parameterization of Force Fields

- Bonded terms: spectroscopy or quantum mechanics
- Lennard-Jones: Small molecular crystals
- Electrostatic: quantum mechanics (fit monopoles to electrostatic potential)
- Many challenges in practice
  - which (model) molecules: availability, representative or not
  - how many atomic classes: transferability and tractability
  - Which properties to parameterize for?
  - Correlation of parameters
  - Higher order/new terms or not?
  - Electron polarization, non-additivity, …
  - water, water, and water
- At the end, do they add up? (cancellation of errors)

Energy Minimization

- Minimization follows gradient of potential to identify stable points on energy surface
  - Let $V(x) = k\,(x-x_0)^2$
  - Begin at x', how do we find x_0 if we don't know V(x) in detail?
    - How can we move from x' to x_0?
  - Steepest descent (SD):
    - $x' \rightarrow x' = x+\delta$
    - $\delta = -dx\,\partial V(x)/\partial x = -dx\,k(x-x_0)$
  - This moves us, depending on the step size dx, toward x_0.
  - On a simple harmonic surface, we will reach the minimum, x_0, i.e. converge, in a certain number of steps related to dx.



Molecular Dynamics

- Objective: $\{r_1(t), \ldots, r_N(t)\} \rightarrow \{r_1(t+\Delta t), \ldots, r_N(t+\Delta t)\}$ f = ma
- Basic idea: solve Newton's equation of motion numerically
  - Given current coordinates (x), velocities (v)
    - Forces can be calculated based on coordinates (from $f = -\partial V/\partial x$)
    - $x(t+\Delta t) = x(t) + v(t)\,\Delta t$
    - $v(t+\Delta t) = v(t) + f(t)/m\,\Delta t$
    - Repeat above operations
- More accurate integrators (better energy conservation)
  - Verlet Algorithm (Verlet J. Chem. Phys. 1967)

consider Taylor's expansions:

$x(t\pm\Delta t) = x(t) \pm v(t)\,\Delta t + 1/2m\,f(t)\,\Delta t^2 \pm 1/6\,d^3x/dt^3\,\Delta t^3 + O(\Delta t^4)$

Adding expansion x(t+Δt) and x(t-Δt) and rearrange:

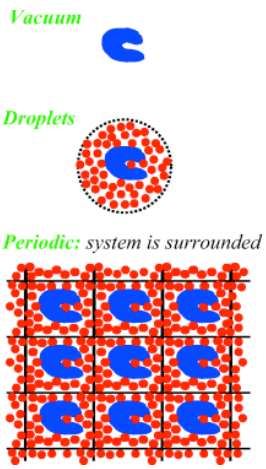
$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + f(t)/m\,\Delta t^2 + O(\Delta t^4)$

velocities lag

Subtracting expansion x(t+Δt) and x(t-Δt) and rearrange:

$v(t) = [x(t+\Delta t) - x(t-\Delta t)/(2\Delta t) + O(\Delta t^3)$

Solvent and Periodic Boundary Conditions



- Surface effects (surface tension)
- No dielectric screening

- Still surface effects (at water – vacuum interface)
- Only partial dielectric screening
- Evaporation of the solvent

Advantage:
- No surface effects

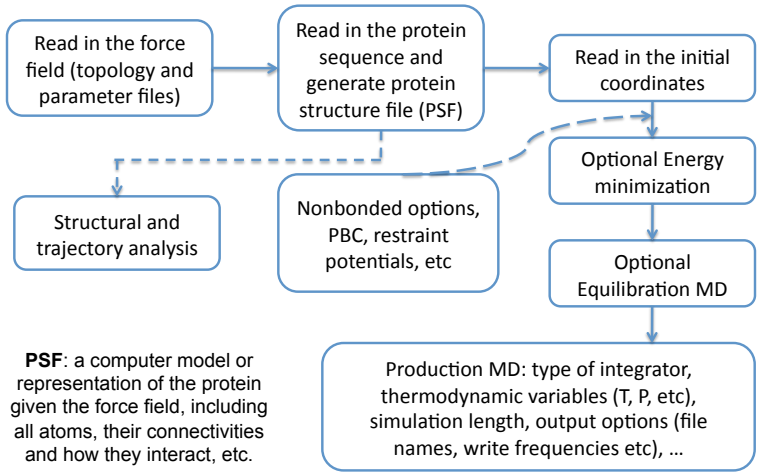
Disadvantage:
- Artificial periodicity
- High effective concentration

van Gunsteren Angew Chem Int Ed (2006)

Controlling Thermodynamic Variables

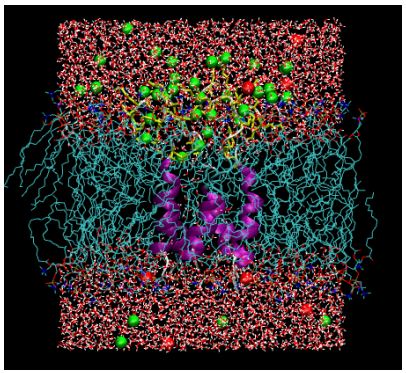
- MD generate statistical ensembles that connect microscopic details to macroscopic/thermodynamic properties
- NVE (microcanonical - Entropy rules!)
- NVT (Canonical - Helmholtz free energy is relevant, A)
  - temperature $T = \sum m\langle v2\rangle/(3k_B)$
- NPT (Isothermal-isobaric - Gibbs free energy is relevant, G)
  - P=kinetic + virial contributions
- Thermostats, barostats, etc., allow one to choose appropriate ensembles
  - Following Nose', Hoover, Evans and others…
  - See Brooks, Curr. Opin. Struct. Biol., 5, 211(1995)]

Basic Flow of a MD Simulation



PSF: a computer model or representation of the protein given the force field, including all atoms, their connectivities and how they interact, etc.

Why is MD so slow?



Channel-forming peptides
in a fully solvated membrane bilayer;
Channel: 1795 atoms; All: 26254 atoms

Simulated Time

1 ns (10^{-9} s)
(500,000 MD steps)

CPU Time

~200 hours (10^6 s)

Wall Time

~1 days (10^5 s) / 8 CPUs

- very small time step required
  - δt ~ fs (10^{-15} s)
- interactions between thousands of atoms need to be computed

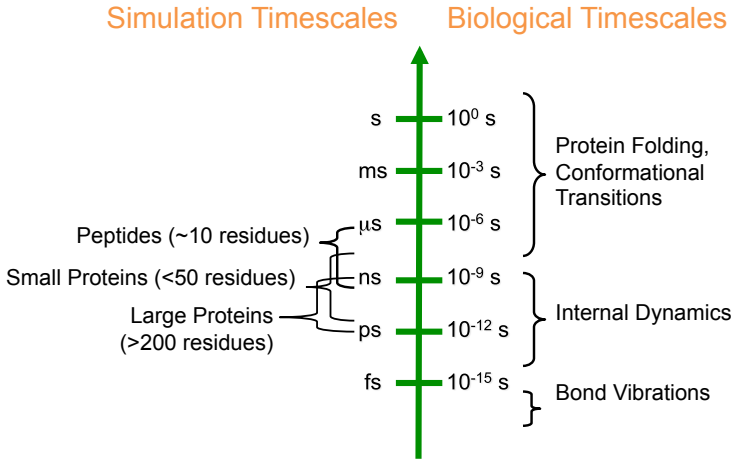
Biological Time Scale

- Bond vibrations 1 fs (10^{-15} s)
- Sugar repuckering 1 ps (10^{-12} s)
- DNA bending 1 ns (10^{-9} s)
- Domain movement 1 ms (10^{-6} s)
- Base pair opening 1 ms (10^{-3} s)

- Transcription 2.5 ms / nucleotide
- Protein synthesis 6.5 ms / amino acid
- Protein folding ~ 10 s (speed limit: μs)
- RNA lifetime ~ 300 s

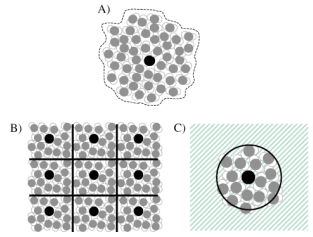
Simulation time should exceed the time scale of interest by ~10-fold !

Gap in Timescales

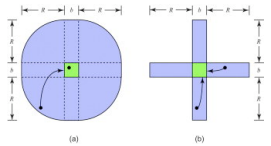


Practical Considerations

- Long-range forces
  - Using cut-off to reduce the number of nonbonded atom pairs (~ 12-15Å)
  - Electrostatic decays slowing (1/r) and cut off does not work well; Particle Mesh Eward (PME) is needed.
- Parallel execution
  - Partition various regions of the system to different CPUs
  - Need to communicate information between nodes; this is a bottleneck
- Simplifications of the model
- Enhanced sampling techniques



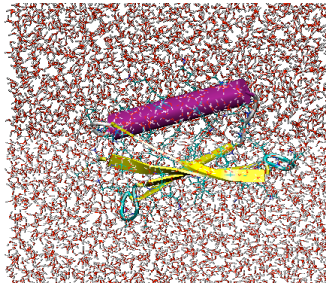
van Gunsteren Angew Chem Int Ed (2006)



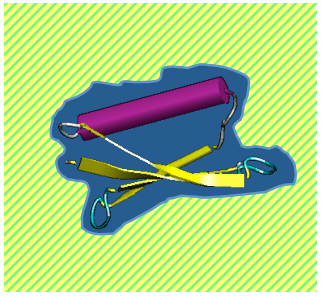
Shaw JCC (2005)

Implicit Solvent

- Solvent increases the system size about 10-fold
- It is possible to describe the mean influence of water w/o explicitly including water



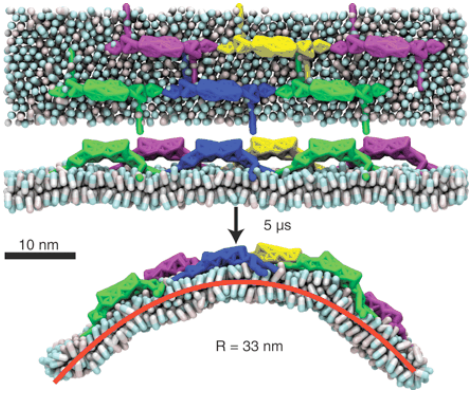
Explicit solvent
Protein: 56 residues (855 atoms)
Solvent: 5411 waters (16233 atoms)



Implicit solvent
Hybrid macroscopic (solvent) / microscopic (solute)

Coarse-Grained Models

- Rely on reduced representation and/or simplified interaction schemes to access larger length and time scales

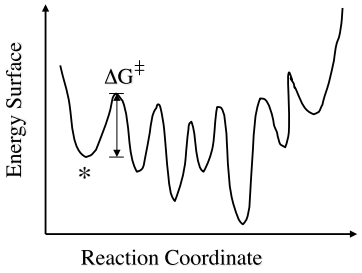


Klein and Shinoda, Science (2008)

Biomembrane sculpting by protein-BAR domains

The simulation shown in the figure was carried out using a box with dimensions 100 x 16 x 50 nm and would correspond to a system of 10 million atoms. Using a shape-based CG model reduces the size to 3265 CG beads. The simulation showed that a concerted action of BAR domains arranged in a lattice results in the development of a global membrane curvature on a time scale of several μs, with the resulting curvature radius of ~30 nm that was observed experimentally.

Barriers, Temperature and Timescales



$$\tau = \tau_0 \exp(\Delta G^{\ddagger}/kT)$$

$\tau_0 \sim 10^{-12}$ s ~ ps

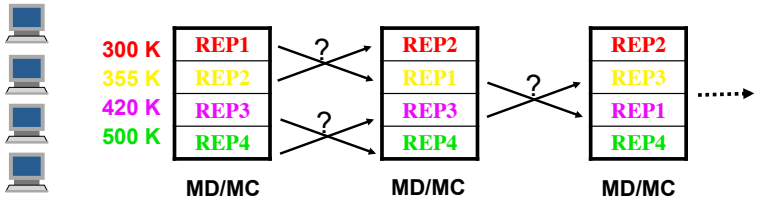
T = 300 K

$\Delta G^{\ddagger}$: 1 kcal/mol, $\tau \sim$ ps
5 kcal/mol, $\tau \sim$ ns
10 kcal/mol, $\tau \sim$ μs

Protein energy landscape is highly complex and rugged with numerous local minima.

Enhanced Sampling Techniques

Replica Exchange (REX)

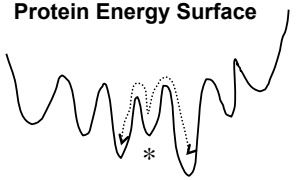


Exchange criteria

$$P_{i \leftrightarrow j} = \begin{cases} 1 & \Delta \le 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases}$$

$$\Delta = (E_i - E_j) \bullet (1/kT_j - 1/kT_i)$$

Protein Energy Surface

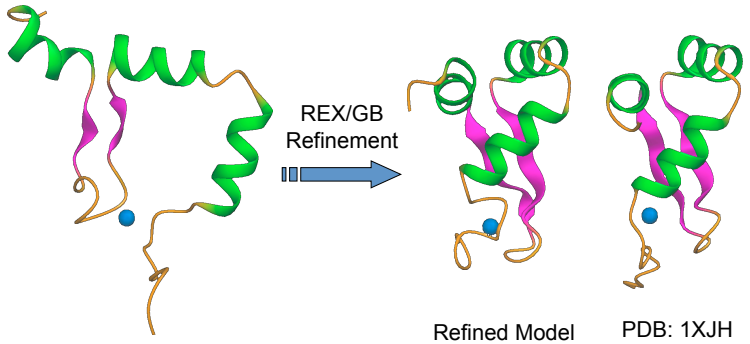


Sugita and Okamoto, *CPL* (1999); MMTSB Tool Set: *http://mmtsb.scripps.edu*

Applications of Modeling

- Main advantages
  - Offer atomistic spatial resolution and femtosecond time resolution
  - Allow probing the system in many nontrivial ways that are not possible or too dangerous experimentally
  - Often much cheaper than doing the experiment itself
  - Can be applied at very large scales (computers are cheap)
  - Can provide theoretical frameworks for experimental studies
- A few prototypical areas
  - Protein structure prediction and calculation
  - Virtual screening and rational drug design
  - Simulation of important systems: mechanisms
  - Interpretation of (static) experimental data
  - Protein misfolding and aggregation
  - Biomolecular engineering: design of new enzymes etc
  - …

NMR Structure Refinement

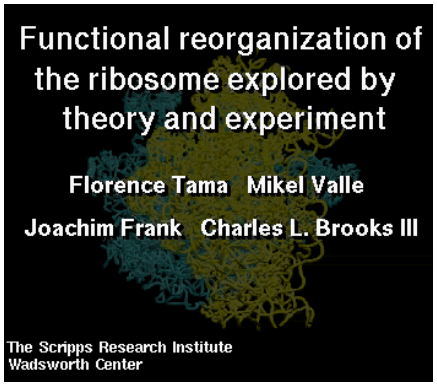


Chen et. al., *JACS* (2004); Chen et al., *J. Biomol NMR* (2004).

Functional Motions of Ribosome

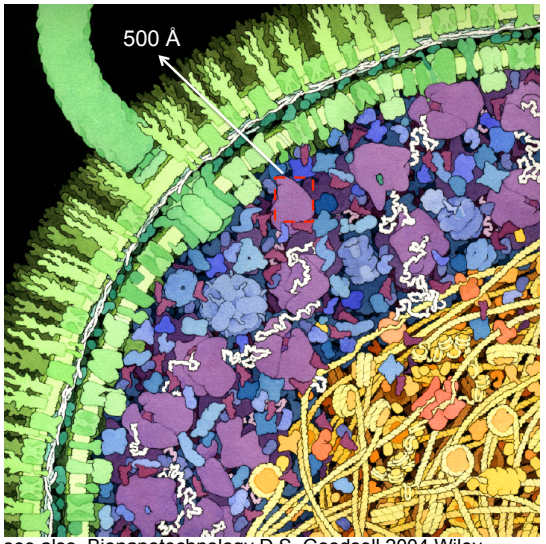
"This movie depicts a ratchet-like rearrangement of the 70S ribosome. The rotation of the 30S ribosomal subunit relative to the 50S subunit shows high correspondence to motion captured in cryo-EM maps of the ribosome and postulated to be a key mechanical step in the translocation of the mRNA•tRNAs complex. "

http://brooks.chem.lsa.umich.edu/



Tama et al. Proc Nat Acad Sci, **100** 9319 (2003).

What do we need to consider?



see also, Bionanotechnology D.S. Goodsell 2004 Wiley