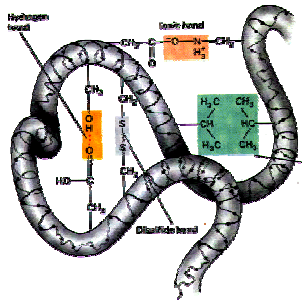
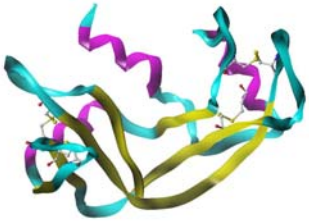
## Tertiary Structure of Protein



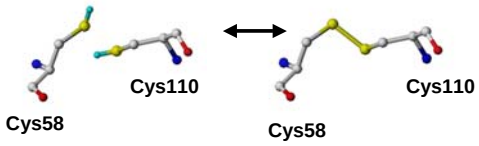
## Anfinsen's experiments, late 1950's through 1960's

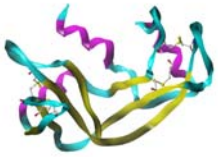


Ribonuclease, an enzyme involved in cleavage of nucleic acids. Structure has a combination of α and β segments and four disulfide bridges

## What are Disulfide Bridges?

Cys-SH + Cys-SH $\underset{\text{Reduction}}{\overset{\text{Oxidation}}{\rightleftharpoons}}$ Cys-S—S-Cys

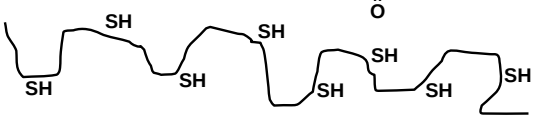


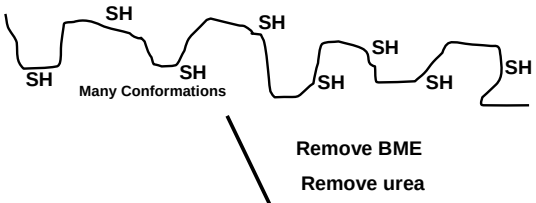
Active, native structure

BME is reducing agent
Urea unfolds proteins

add β-mercaptoethanol (BME)
$HS\text{-}CH_2\text{-}CH_2\text{-}OH$
add urea, $H_2N\text{-}\underset{\underset{O}{\|}}{C}\text{-}NH_2$

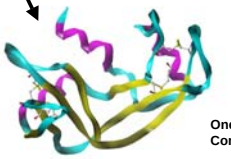



Denatured, inactive, "random coil", many conformations

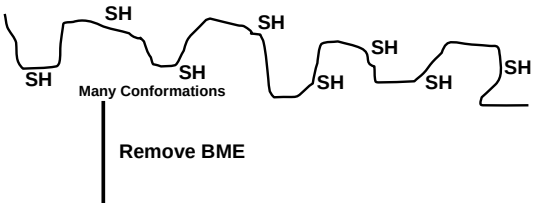



Remove BME
Remove urea

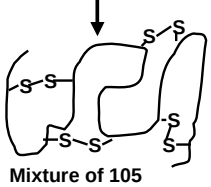
- Native structure
- fully active
- 4 disulfide bond correct



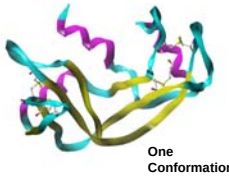
One Conformation




Remove BME

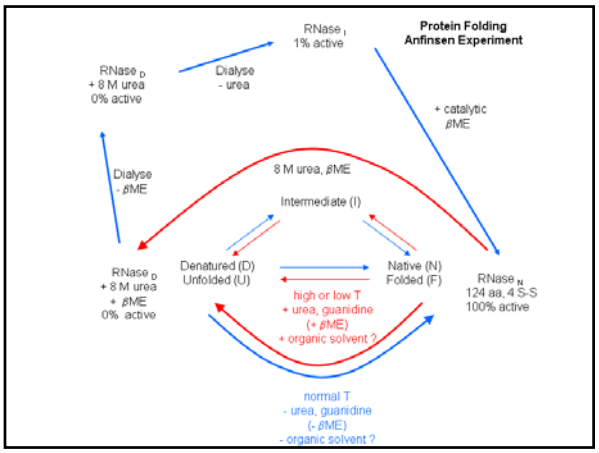


Remove urea



One Conformation

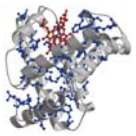
Mixture of 105 different conformations, 1% active




## The Protein Folding Problem

Levinthal's paradox – Consider a 100 residue protein.
If each residue can take only 3 positions,
there are $3^{100} = 5 \times 10^{47}$ possible conformations.

If it takes $10^{-13}$s to convert from 1 structure to another,
exhaustive search would take $1.6 \times 10^{27}$ years!

MACGT... ? 

*"Given a particular sequence of amino acid residues (primary structure), what will the tertiary/quaternary structure of the resulting protein be?"*

## Protein Structure Prediction and Protein Folding

Fundamental Questions

**Protein Structure Prediction**

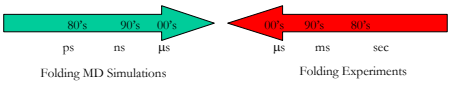
- What is the structure of this protein?
  - Can be experimentally determined, today we know the structure of ~35,000 proteins
  - Can be predicted for some proteins, usually in ~1 day on today's computers

**Protein Folding**

- How does this protein form this structure?
  - The process or mechanism of folding
  - Limited experimental characterization
- Why does this protein form ***this*** structure?
  - Why not some other fold?
  - Why so quickly? -> Levinthal's Paradox: As there are an astronomical number of conformations possible, an unbiased search would take too long for a protein to fold. Yet most proteins fold in less than a second!

## Protein Folding: Fast Folders

Time Scale:




- Trp-cage, designed mini-protein (20 aa): 4µs
- β-hairpin of C-terminus of protein G (16 aa) : 6µs
- Engrailed homeodomain (En-HD) (61 aa): ~27µs
- WW domains (38-44 aa): >24µs
- Fe(II) cytochrome $b_{562}$ (106 aa): extrapolated ~5µs
- B domain of protein A (58 aa): extrapolated ~8µs

## Structure Prediction Methods

```
1  QQYTA KIKGR
11 TFRNE KELRD
21 FIEKF KGR
```

Algorithm 

- **Secondary structure (only sequence)**
- **Homology modeling (using related structure)**
- **Fold recognition**
- ***Ab-initio* 3D prediction**

## Homology Modeling

- **Assumes similar (homologous) sequences have very similar tertiary structures**
- **Basic structural framework is often the same (same secondary structure elements packed in the same way)**
- **Loop regions differ**
  - **Wide differences possible, even among closely related proteins**

## Threading

- Given:
  - sequence of protein P with unknown structure
  - Database of known folds
- Find:
  - Most plausible fold for P
  - Evaluate quality of such arrangement
- Places the residues of unknown P along the backbone of a known structure and determine stability of side chains in that arrangement

## Strategies for Protein Structure Prediction

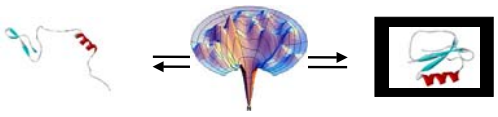
| | Comparative Modeling | Fold Recognition | Ab Initio |
|---|---|---|---|
| Method | 1. Identify sequence homologs as templates<br>2. Use sequence alignment to generate model<br>3. Fill in unaligned regions<br>4. Improves with data | 1. Fold classification<br>2 3D-Profiles<br>3. Improves with data | 1. Representation<br>2. Force field<br>3. Global Optimization<br>4. Structure at global minimum<br>5. Can discover new folds |
| Drawbacks | 1. Requires > 25% sequence identity<br>2. Loops and sidechain conformations are critical | 1. Needs good number of proteins in each fold<br>2. Critically dependent on scoring function | 1. Computationally intensive<br>2. Physical modeling |
| Resolution | < 3 A | 3 - 7 A | > 5 A |
| Time to Compute | < Day | ~ Day | >> Day |

## Complementarity of the Methods

- **X-ray crystallography- highest resolution structures; faster than NMR**
- **NMR- enables widely varying solution conditions; characterization of motions and dynamic, weakly interacting systems**
- **Computation- fundamental understanding of structure, dynamics and interactions; models without experiment; very fast**

The protein sequence contains all information needed to create a correctly folded protein.

- Many proteins fold spontaneously to their native structure
- Protein folding is relatively fast
- Chaperones speed up folding, but do not alter the structure



## Forces driving protein folding

- It is believed that *hydrophobic collapse* is a key driving force for protein folding
  - Hydrophobic core
  - Polar surface interacting with solvent
- Minimum volume (no cavities)
- Disulfide bond formation stabilizes
- Hydrogen bonds
- Polar and electrostatic interactions

Native state is typically only 5 to 10 kcal/mole more stable than the unfolded form

**Four models that could account for the rapid rate of protein folding during biological protein synthesis.**

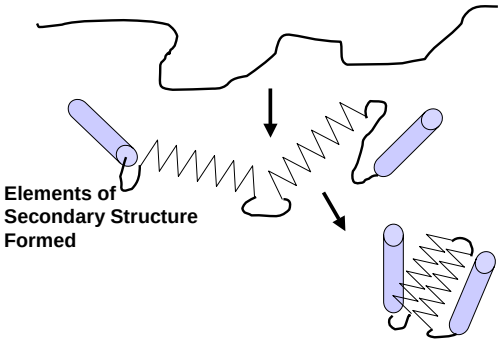
**- The Framework Model**

**- The Nucleation Model**

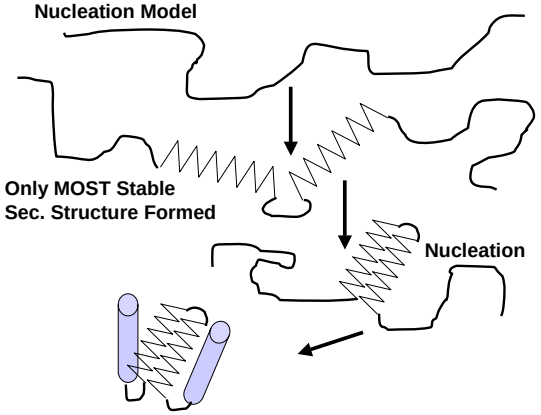
**- The "Molten Globule" Model**

**- "Folding Funnels"**

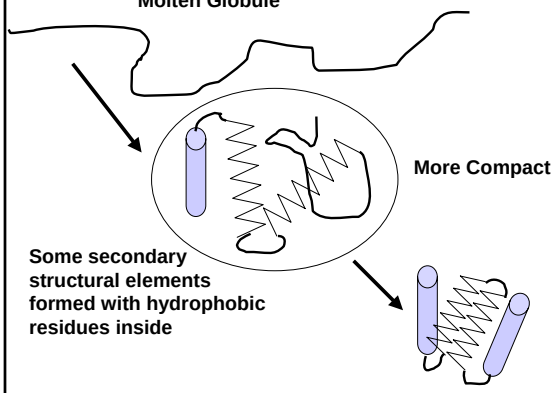
## Framework Model



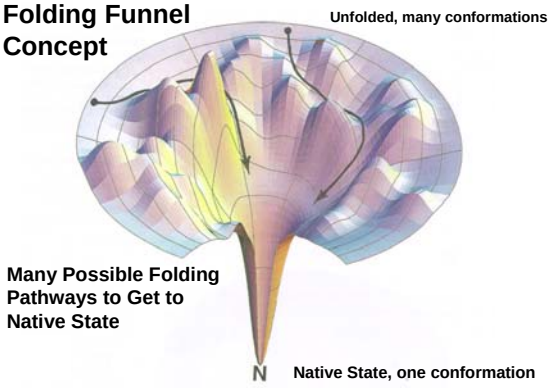

## Nucleation Model




## Molten Globule

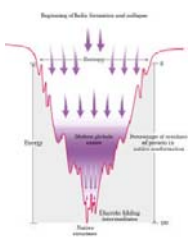



## Folding Funnel Concept



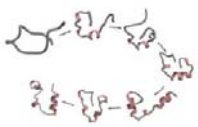

## Thermodynamics of Protein Folding

- Bond stretching: $10^{-14}$ - $10^{-13}$ sec.
- Elastic vibrations: $10^{-12}$ - $10^{-11}$ sec.
- Rotations of surface sidechains: $10^{-11}$ - $10^{-10}$ sec.
- Hinge bending: $10^{-11}$ - $10^{-7}$ sec.
- Rotation of buried side chains: $10^{-4}$ - 1 sec.
- Protein folding: $10^{-6}$ - $10^{2}$ sec.



Free Energy Funnel

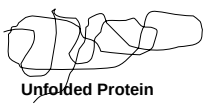
Simulated folding in 1 μsec;
peptide in a box of water



## Entropy and Enthalpy in Protein Folding

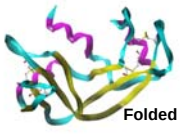
$$\Delta G = \Delta H - T\Delta S$$

bonding flexibility



Unfolded Protein

ΔH, small, negative
ΔS, large, positive



Folded Protein

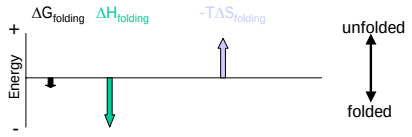
ΔH, large, negative
ΔS, small, positive

Compensation in entropy and enthalpy for protein
Contribution of entropy of water molecules released upon folding
ΔS of water is large and positive

## Thermodynamics of Protein Folding

$$\Delta G_{folding} = G_{folded} - G_{unfolded} =$$

$$(H_{folded} - H_{unfolded}) - T(S_{folded} - S_{unfolded}) = \Delta H_{folding} - T\Delta S_{folding}$$



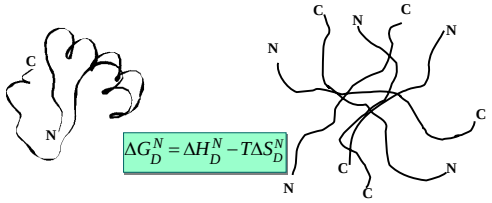

Folded proteins are highly ordered

$\therefore \Delta S_{folding}$ negative, so $-T\Delta S_{folding}$ is a positive quantity

$\Delta H_{folding}$ is a negative quantity - enthalpy is favored in folded state.

Total Gibbs free energy difference is negative – folded state favoured

## Native state (N) / Denatured state




$$\Delta G_D^N = \Delta H_D^N - T\Delta S_D^N$$

Size of cavity in solvent: ~6500Å²

Average size of cavity in solvent: 20,500Å²

ΔS chain: **significantly decreased, due to the well defined conformation**

ΔS chain: **large, due to the large number of different conformations**

Non-bonded interactions: **intra-molecular**

Non-bonded interactions: **inter-molecular**

Compact structure

Non compact structure

## Factors that disrupt the Native state

1) **ELECTROLYTE ADDITION**
   - **interference with the colloid state**

2) **INSOLUBLE SALT FORMATION**
   - **Protein+Trichloracetate**

3) **ORGANIC SOLVENTS**
   - **ETHANOL - interferes with the dielectric constant**

4) **HEAT DENATURATION**
   - **more energy in system (bonds break)**

5) **pH**
   - **destroys charge**
   - **destroys ability to interact with water**

6) **DESTRUCTION OF HYDROGEN BONDING**
   - **UREA - known H-bond disrupter**

## Thermodynamic Description of Protein Folding

The native and unfolded states are in equilibrium, the folding reaction can be quantified in terms of thermodynamics.

The equilibrium ($N \leftrightarrow U$) between the native (N) and unfolded (U) states is defined by the equilibrium constant, K, as:

$$K = [U]/[N] = K_U$$

The difference in Gibbs free energy ($\Delta G$) between the unfolded and native states is then:

$$\Delta G = -RT \ln K$$

For $K_u$, a positive $\Delta G$ indicates that the native state is more stable.

The free energy is composed of both enthalpic and entropic contributions:

$$\Delta G = \Delta H - T\,\Delta S$$

where $\Delta H$ and $\Delta S$ are the enthalpy and entropy change, respectively, upon unfolding.

**Thermal Unfolding**

Since $\Delta H$ and $\Delta S$ are strongly temperature-dependent, $\Delta G$ is better expressed as:

$$\Delta G = \Delta H_1 + \Delta C_P\,(T-T_1) - T\,[\,\Delta S_1 + \Delta C_P \ln(T/T_1)]$$

where the subscript "1" indicates the value of $\Delta H$ and $\Delta S$ at a reference temperature, $T_1$, and $\Delta C_p$ is the specific heat or heat-capacity change.

Most proteins denature reversibly allowing thermodynamic analysis.


## Factors that disrupt the Native state

1) **ELECTROLYTE ADDITION**
   - **interference with the colloid state**

2) **INSOLUBLE SALT FORMATION**
   - **Protein+Trichloracetate**

3) **ORGANIC SOLVENTS**
   - **ETHANOL - interferes with the dielectric constant**

4) **HEAT DENATURATION**
   - **more energy in system (bonds break)**

5) **pH**
   - **destroys charge**
   - **destroys ability to interact with water**

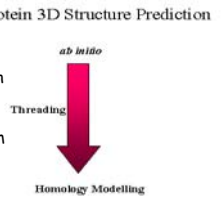
6) **DESTRUCTION OF HYDROGEN BONDING**
   - **UREA - known H-bond disrupter**

## Solving Protein Structures

Only 2 kinds of techniques allow one to get atomic resolution pictures of macromolecules

- X-ray Crystallography (first applied in 1961 - Kendrew & Perutz)
- NMR Spectroscopy (first applied in 1983 - Ernst & Wuthrich)

- Structure ⟷ Function
- Structure ⟷ Mechanism
- Structure ⟷ Origins/Evolution
- Structure-based Drug Design
- Solving the Protein Folding Problem




## Ab Initio Prediction

- Predicting the 3D structure without any "prior knowledge"
- Used when homology modelling or threading have failed (no homologues are evident)
- Equivalent to solving the "Protein Folding Problem"
- Still a research problem

## Ab Initio Folding

Two Central Problems

- Sampling conformational space ($10^{100}$)
- The energy minimum problem

The Sampling Problem (Solutions)

- Lattice models, off-lattice models, simplified chain methods

The Energy Problem (Solutions)

- Threading energies, packing assessment, topology assessment

## Problems in Protein Folding

- Two key questions:
  - Evaluation – how can we tell a correctly-folded protein from an incorrectly folded protein?
    - H-bonds, electrostatics, hydrophobic effect, etc.
    - Derive a function, see how well it does on "real" proteins
  - Optimization – once we get an evaluation function, can we optimize it?
    - Simulated annealing/Monte Carlo

| *Interaction* | *Approx. bond strength in kJ/mole* |
|---|---|
| **Covalent bonds** | **> 200 (ranging up to 900)** |
| **Ionic** | **20-40** |
| **Hydrogen bond** | **~5-20** |
| **Hydrophobic** | **~ 8** |
| **van der Waals** | **~ 4** |

## Evaluation of Protein Folds

- Empirical potential functions
  - Residue-based: spatial relationships among residues
  - Stereochemistry-based: molecular interactions (covalent, electrostatic, etc.) with coefficients
- Ab-initio potential functions
- Procheck, etc.
- Full molecular dynamics
  - Very computationally expensive

AMBER (*Assisted Model Building with Energy Refinement* ) force field

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \sum_{1}^{3} \frac{V_n}{2}[1 + \cos(n\omega)]$$

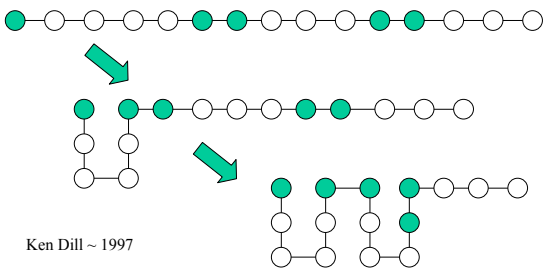
$$+ \sum_{i<j}^{atoms} \left( \frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} \right) + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon r_{ij}}$$

## Polypeptides

- Represented by a range of approaches or approximations including:
  - all atom representations in cartesian space
  - all atom representations in dihedral space
  - simplified atomic versions in dihedral space
  - tube/cylinder/ribbon representations
  - lattice models

## Lattice Models

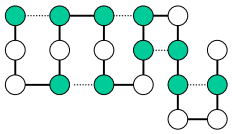
- The "hydrophobic zipper" effect:



Ken Dill ~ 1997

## Scoring Lattice Models

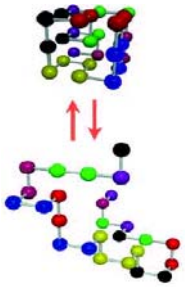
- H/P model scoring: count noncovalent hydrophobic interactions.



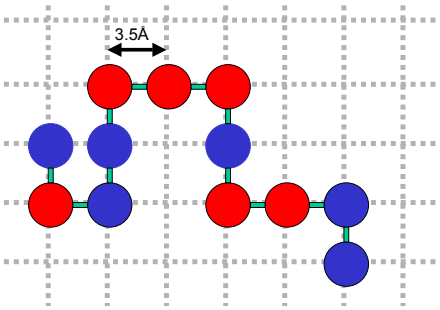
- Sometimes:
  - Penalize for buried polar or surface hydrophobic residues

## Fold Optimization

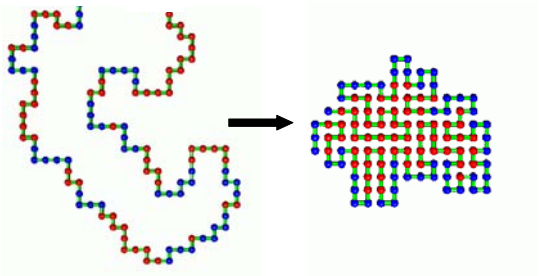
- Simple lattice models (HP-models)
  - Two types of residues: hydrophobic and polar
  - 2-D or 3-D lattice
  - The only force is hydrophobic collapse
  - Score = number of H–H contacts



## A Simple 2D Lattice

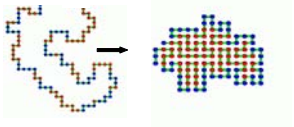



## Lattice Folding



## Lattice Algorithm

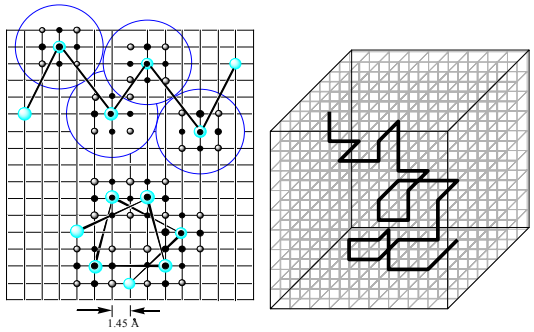
- *Build a "n x m" matrix (a 2D array)*
- *Choose an arbitrary point as your N terminal residue (start residue)*
- *Add or subtract "1" from the x or y position of the start residue*
- *Check to see if the new point (residue) is off the lattice or is already occupied*
- *Evaluate the energy*
- *Go to step 3 and repeat until done*



- *Red = hydrophobic*
- *Blue = hydrophilic*

- *If Red is near empty space E = E+1*
- *If Blue is near empty space E = E-1*
- *If Red is near another Red E = E-1*
- *If Blue is near another Blue E = E+0*
- *If Blue is near Red E = E+0*

## More Complex Lattices



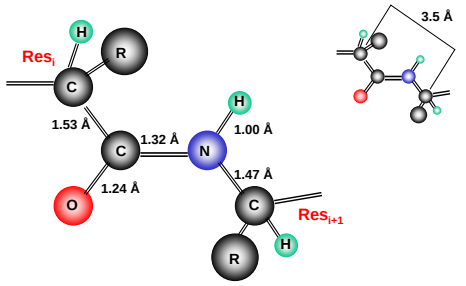

## What can we do with lattice models?

- For smaller polypeptides, exhaustive search can be used
  - Looking at the "best" fold, even in such a simple model, can teach us interesting things about the protein folding process

## More realistic models

- Higher resolution lattices (45° lattice, etc.)
- Off-lattice models
  - Local moves
  - Optimization/search methods and $\phi/\psi$ representations
    - Greedy search
    - Graph theoretical methods
    - Monte Carlo, simulated annealing, etc.

## Non-Lattice Models

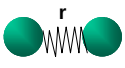



## Non-Lattice Models

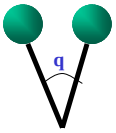
- With a more realistic off-lattice model, we need a better *energy function* to evaluate a conformation (fold).
- Theoretical force field:
  $\Delta G = \Delta G_{van\ der\ Waals} + \Delta G_{h\text{-}bonds} + \Delta G_{solvent} + \Delta G_{coulomb}$
- Empirical force fields

## Energy Terms

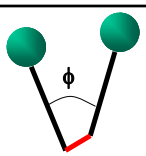
### Covalent




Stretching
$K_r(r_i - r_j)^2$

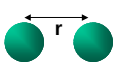



Bending
$K_\theta(\theta_i - \theta_j)^2$

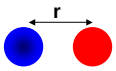



Torsional
$K_\phi(1–\cos(n\phi_j))^2$

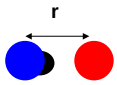
### Noncovalent




van der Waals
$A_{ij}/r^6 - B_{ij}/r^{12}$




Coulomb
$q_iq_j/4\pi\varepsilon r_{ij}$




H-bond
$C_{ij}/r^{10} - D_{ij}/r^{12}$

## Bonding Terms: bond stretch

- Most often Harmonic

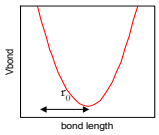
$$V_{bond} = \sum_{bonds} \frac{1}{2} k_r (r - r_0)^2$$

- Morse Potential for dissociation studies

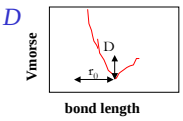
$$V_{Morse} = \sum_{bonds} D[e^{-a(r-r_0)} - 1]^2 - D$$

Two new parameters:
D: dissociation energy
a: width of the potential well

Harmonic Potential



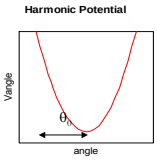

Morse Potential



## Bonding Terms: angle bending

- Most often Harmonic

$$V_{angle} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2$$




- CHARMM force field's Urey-Bradley angle term:

$$V_{UB} = \sum_{UB} \frac{1}{2} k_{UB} (s - s_0)^2$$

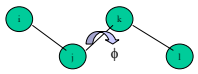
This UB term is only found in CHARMM force field to optimize the fit to vibrational spectra.
s: the 1,3-distance.

Mackerell et al. J. Phys. Chem. B 102, 3586, 1998

## Bonding Terms: Torsions

- Torsion energy: rotation about a bond (dihedral angles)

$$U_{torsion} = \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$




i-j-k-l

Vn: force constant
n: periodicity of the angle ( determines how many peaks and wells in the potential, often from 1-6 )
δ: phase of the angle (often 0° or 180°)

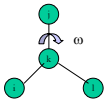
## Bonding Terms: Improper Torsions

- Improper torsion is not a regular torsion angle. It is used to describe the energy of out-of-plane motions. It is often necessary for planar groups, such as sp2 hybridized carbons in carbonyl groups and in aromatic rings, because the normal torsion terms described above is not sufficient to maintain the planarity (ω~0).

$$U_{improper} = \sum_{improper} \frac{V_2}{2} [1 + \cos(2\omega - 180^\circ)]$$

or

$$U_{improper} = \sum_{improper} \frac{k_w}{2} (\omega - \omega_0)^2$$

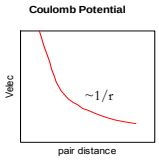



i-j-k-l

## Non-bonded Terms

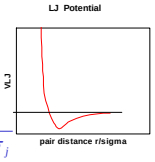
- Electrostatic interactions (Coulomb's Law)

$$V_{elec} = \frac{1}{4\pi\varepsilon} \sum_{i<j} \frac{q_i q_j}{r_{ij}}$$




- Lennard-Jones interactions

$$V_{LJ} = \sum_{i<j} 4\varepsilon_{ij} \left[ \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right]$$

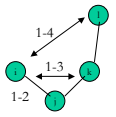



- Combination Rules for LJ

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \quad \sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \quad \sigma_{ij} = \sqrt{\sigma_i \sigma_j}$$

## 1-4 Non-bonded Interactions

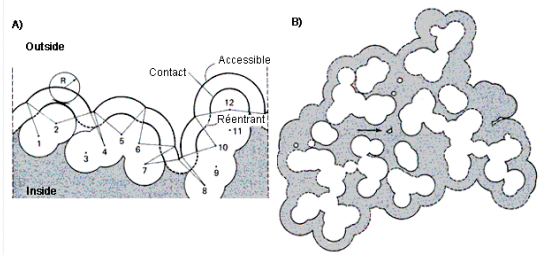
- Non-bonded exclusions
  - 1-2 and 1-3 interactions excluded
  - 1-4 interactions partially excluded
- 1-4 interaction scalings
  - OPLSAA scales by 0.5 for both electrostatic and LJ
  - AMBER94 scales 0.5 for LJ and 1/1.2 for electrostatic interaction
  - CHARMM22 has special 1,4-terms



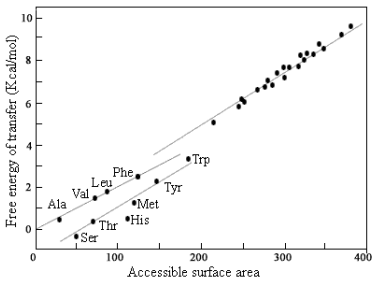

Even though they are non-bonded interactions, 1-4 terms are often calculated along with bonded terms.

## The hydrophobic effect

The free energy gain from burying a hydrophobic group is proportional to the surface area buried



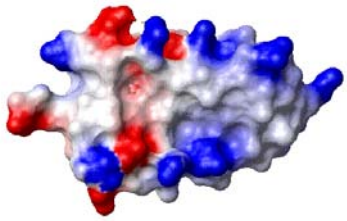
Linear relation between the solvent accessible surface area and the transfer free energy of amino acids




$\Delta G_{transfer} = -\gamma ASA$

$\gamma = 0.025\ cal/Å^2$

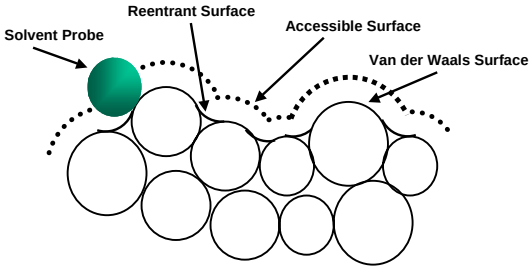
## Accessible Surface Area



```
QHTAWCLTSEQHTAAVIWDCETPGKQNGAYQEDCA
HHHHHHCCEEEEEEEEEECCHHHHHHHCCCCCCC
```

```
QHTAWCLTSEQHTAAVIWDCETPGKQNGAYQEDCAMD
BBPPBEEEEEPBPBPBPBBPEEEPBPEPEEEEEEEEE
10562987994152515104789414969899999999
```

## Accessible Surface Area




## Accessible Surface Area Calculations

- DSSP - Database of Secondary Structures for Proteins (swift.embl-heidelberg.de/dssp)
- Connolly Molecular Surface Home Page
  - http://www.biohedron.com/
- Naccess Home Page
  - http://sjh.bi.umist.ac.uk/naccess.html
- ASA Parallelization
  - http://cmag.cit.nih.gov/Asa.htm
- Protein Structure Database
  - http://www.psc.edu/biomed/pages/research/PSdb/

## Force Fields: Typical Energy Functions

| | |
|---|---|
| $U = \sum_{bonds} \frac{1}{2} k_r (r - r_0)^2$ | Bond stretches |
| $+ \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2$ | Angle bending |
| $+ \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$ | Torsional rotation |
| $+ \sum_{improper} V(improper\ torsion)$ | Improper torsion (sp2) |
| $+ \sum_{elec} \frac{q_i q_j}{r_{ij}}$ | Electrostatic interaction |
| $+ \sum_{LJ} [\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}]$ | Lennard-Jones interaction |

## Which Force Field to Use?

- Most popular force fields: CHARMM, AMBER and OPLSAA
- OPLSAA(2000): Probably the best available force field for condensed-phase simulation of peptides. Work to develop parameterization that will include broader classes of drug-like molecules is ongoing. GB/SA solvation energies are good.
- MMFF: An excellent force field for biopolymers and many drug-like organic molecules that do not have parameters in other force fields.
- AMBER*/OPLS*: Good force fields for biopolymers and carbohydrates; many parameters were added in MacroModel which extend the scope of this force field to a number of important organic functional groups. GB/SA solvation energies range from moderate (AMBER*) to good (OPLS*).
- AMBER94: An excellent force field for proteins and nucleic acids. However, there are no extensions for non-standard residues or organic molecules, also there is a alpha-helix tendency for proteins. AMBER99 fixes this helix problem to some degree, but not completely.
- MM2*/MM3*: Excellent force fields for hydrocarbons and molecules with single or remotely spaced functional groups. GB/SA solvation energies tend to be poor relative to those calculated with other force fields.
- CHARMM22: Good general purpose force field for proteins and nucleic acids. A bit weak for drug-like organic molecules.
- GROMOS96: Good general purpose force field for proteins, particularly good for free energy perturbations due to soft-core potentials. Weak for reproducing solvation free energies of organic molecules and small peptides.

http://www.schrodinger.com/docs/mm7.1/html/faqs/which_ffield.html

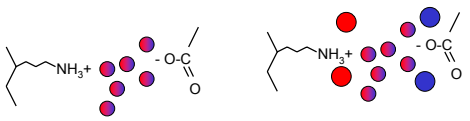
## Force Field Parameterization

- Equilibrium bond distances and angles: X-ray crystallography
- Bond and angle force constants: vibrational spectra, normal mode calculations with QM
- Dihedral angle parameters: difficult to measure directly experimentally; fit to QM calculations for rotations around a bond with other motions fixed
- Atom charges: fit to experimental liquid properties, ESP charge fitting to reproduce electrostatic potentials of high level QM, X-ray crystallographic electron density
- Lennard-Jones parameters: often most difficult to determine, fit to experimental liquid properties, intermolecular energy fitting

## Applications

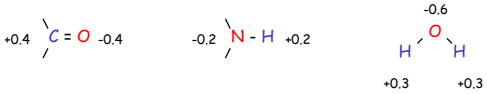
- NMR or X-ray structure refinement
- Protein structure prediction
- Protein folding kinetics and mechanics
- Conformational dynamics
- Global optimization
- DNA/RNA simulations
- Membrane proteins/lipid layers simulations

Dielectric constant



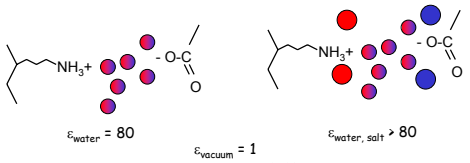

$\varepsilon_{water} = 80$ $\varepsilon_{vacuum} = 1$ $\varepsilon_{protein\ interior} = 2\text{-}10$ $\varepsilon_{water,\ salt} > 80$

Partial Charges

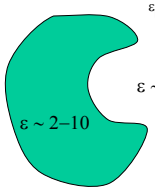



If you know the position of every partial charge (including water), you do not need a dielectric constant.

Dielectric constant



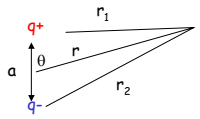

$\varepsilon_{water} = 80$ $\varepsilon_{vacuum} = 1$ $\varepsilon_{protein\ interior} = 2\text{-}10$ $\varepsilon_{water,\ salt} > 80$




The electrostatic potential at any point relative to fixed known charges even in the presence of mobile charges using the Poisson - Boltzmann Equation.

Dipole - Monopole Interactions

$U = \Sigma U(n) = U_1 + U_2$




$U = 1/4\pi\varepsilon_o\,(q/r_1 - q/r_2) = q/4\pi\varepsilon_o\,(r_1 - r_2/\ r_1r_2)$

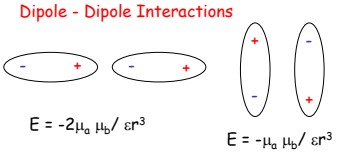
if $r \gg a$ then $r_2 - r_1 \sim a\cos\theta$ and $r_1r_2 = r^2$

$U = qa/4\pi\varepsilon o\ (\cos\theta/\ r^2)$

dipole

U is a function of θ and r. If you rotate around the dipole axis, there is no change in the value of U

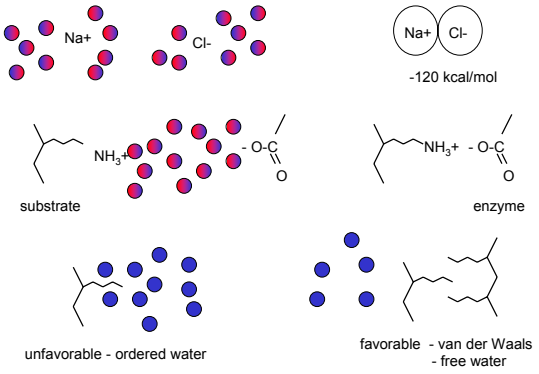
Dipole - Dipole Interactions




$E = -2\mu_a\,\mu_b/\ \varepsilon r^3$ $E = -\mu_a\,\mu_b/\ \varepsilon r^3$

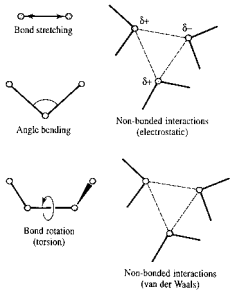
μ = dipole moment = Zd
water = 1.85 D
peptide bond = 3.5 D
retinal = 15 D

Interaction energy is dependent on orientation and distance




hydrophobic effect = ~50 cal/mol/Å²

## Empirical Force Fields and Molecular Mechanics




- describe interaction of atoms or groups
- the parameters are "empirical", i.e. they are dependent on others and have no direct intrinsic meaning
- Examples:
  GROMOS96 (van Gusteren)
  CHARMM (M. Karplus)
  AMBER (Kollman)

## Example for a (very) simple Force Field:

$$\nu = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_N}{2}\left(1+\cos\left(n\omega - \gamma\right)\right) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\pi\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$

Complete Energy Function:

$$H = \sum_{atoms}\frac{p^2}{2m} + \sum_{bond-stretch}\frac{1}{2}k_r(r-r_{eq})^2 + \sum_{bond-angle-bending}\frac{1}{2}k_\theta(\theta-\theta_{eq})^2 + \sum_{bond-rotation}\frac{v_n}{2}[1+\cos(n\phi-\gamma)] + \sum_{S-bond}[V_0(1-e^{-a(r-r_0')})^2 - V_0] + \sum_{H-bond}[V_0(1-e^{-a(r-r_0')})^2 - V_0] + \sum_{non-bonded}[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_{ij} r_{ij}}]$$

AMBER (*Assisted Model Building with Energy Refinement* ) force field

$$E_{total} = \sum_{bonds} K_r(r-r_{eq})^2 + \sum_{angles} K_\theta(\theta-\theta_{eq})^2 + \sum_{dihedrals}\sum_{1}^{3}\frac{V_n}{2}[1+\cos(n\omega)] + \sum_{i<j}^{atoms}\left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}}\right) + \sum_{i<j}^{atoms}\frac{q_i q_j}{\varepsilon r_{ij}}$$

Sources of force parameters:

Bonds, VdW, Electrostatic (for amino acids, nucleotides only):
- AMBER: J. Am. Chem. Soc. 117, 5179-5197
- CHARMM: J. Comp. Chem. 4, 187-217

H-bonds (Morse potential):
- Nucleic Acids Res. 20, 415-419.
- Biophys. J. 66, 820-826

Electrostatic parameters of organic molecules need to be computed individually by using special software (such as Gaussian)

**Average energy scale for various interactions:**

| Energy Term | Scale (kcal/mol) |
|---|---|
| Bond stretching | 100 |
| Angle Bending | 10 |
| Torsion | 1 |
| Hydrogen Bond | 2 |
| Electrostatic interaction | 0.5 |
| Van der Waals | 0.1 |

Concept of energy scale is Important for molecular modeling

**Average bond energies in units of kJ/mol (1kJ/mol=0.239 kcal/mol):**

A. Single bonds:

| | S | P | O | N | C | H |
|---|---|---|---|---|---|---|
| **H** | 339 | 318 | 463 | 389 | 414 | 436 |
| **C** | 259 | 264 | 351 | 293 | 347 | |
| **N** | | 209 | 201 | 159 | | |
| **O** | | 351 | 138 | | | |
| **P** | 230 | 213 | | | | |
| **S** | 213 | | | | | |

B. Multiple bonds:

| | | | |
|---|---|---|---|
| N=N | 418 | C=C | 611 |
| N≡N | 946 | C≡C | 837 |
| C=N | 615 | C=O (in $CO_2$) | 803 |
| C≡N | 891 | C=O (as in $H_2C=O$) | 745 |
| O=O | 498 | C≡O | 1075 |

Energy Minimization

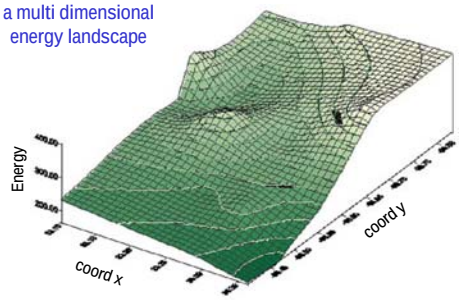
- E = f(x)
- E is a function of coordinates either cartesian or internal
- At minimum the first derivatives are zero and the second derivatives are all positive

$$\frac{dE}{dx_i} = 0$$

$$\frac{d^2E}{dx_i^2} \rangle 0$$

Potential Energy Surface (PES)

a multi dimensional energy landscape




- Systematic Searching
  - explore the whole PES
- Stochastic Searching
  - find "all" low energy minima by generating starting conformation with random changes of rotatable dihedral angles (sometimes combined with random perturbation of the Cartesian coordinates) followed by minimization
- Monte Carlo Simulations
  - generate a Boltzmann distributed ensemble of conformations, can estimate macroscopic thermodynamic properties
- Molecular Dynamics
  - Simulates the time dependent motion of the molecular system, can estimate macroscopic thermodynamic properties
- Simulated Annealing
  - Playing with the temperature (T) in either MD or MC simulations to speed up search for low energy minima
- Distance geometry
  - method for generating conformations that satisfy experimental constraints

Systematic Searching

$$\text{Number of Conformers} = \left(\frac{360}{x}\right)^n$$

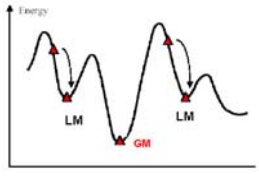
where $x$ is the angle increment
$n$ is the number of rotatable bonds

The exhaustive nature of the search is a *combinatorial problem*

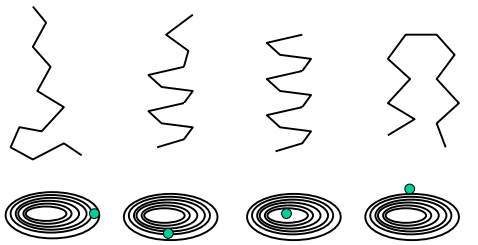
| Rotatable Bonds, $n$ | Increment, $x$ | Conformers |
|---|---|---|
| 3 | 30° | 1728 |
| 3 | 15° | 13824 |
| 3 | 7.5° | 110592 |
| 4 | 30° | 20736 |
| 5 | 30° | 248832 |
| 6 | 30° | 2985984 |

**Molecular Mechanics - Energy Minimization**

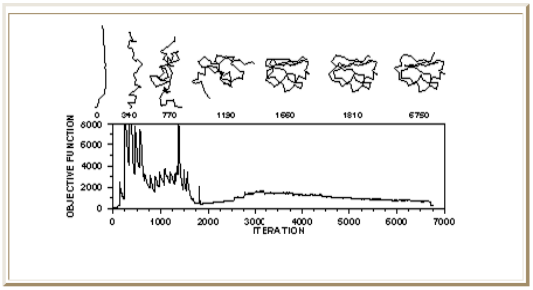
- **The energy of the system is minimized. The system tries to relax**
- **Typically, the system relaxes to a local minimum (LM).**



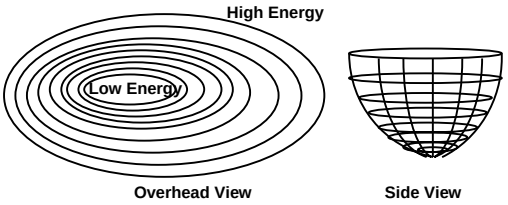
Conformational Sampling



Mid-energy lower energy lowest energy highest energy




- Treat Protein molecule as a set of balls (with mass) connected by rigid rods and springs
- Rods and springs have empirically determined force constants



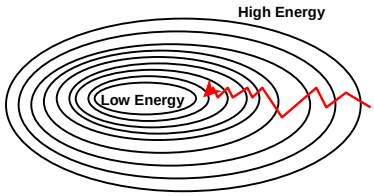

Minimization Methods

- Energy surfaces for proteins are complex hyperdimensional spaces
- Biggest problem is overcoming local minimum problem
- Simple methods (slow) to complex methods (fast)
  - Monte Carlo Method
  - Steepest Descent
  - Conjugate Gradient

**Steepest Descent & Conjugate Gradients**

- Frequently used for energy minimization of large (and small) molecules
- Ideal for calculating minima for complex (i.e. non-linear) surfaces or functions
- Both use derivatives to calculate the slope and direction of the optimization path
- Both require that the scoring or energy function be differentiable (smooth)

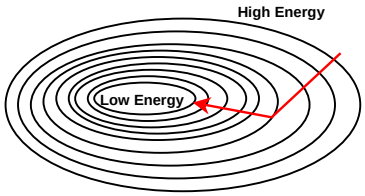
Steepest Descent




**Makes small locally steep moves down gradient**

The steepest descent method uses the first derivative to determine the direction towards the minimum.

Conjugate Gradient Minimization

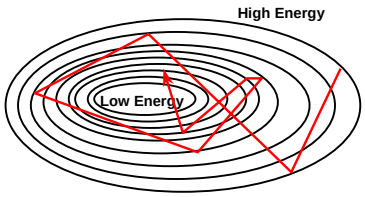



Includes information about the prior history of path

### Monte Carlo Algorithm

- Generate a conformation or alignment (a state)
- Calculate that state's energy or "score"
- If that state's energy is less than the previous state accept that state and go back to step 1
- If that state's energy is greater than the previous state accept it if a randomly chosen number is < $e^{-E/kT}$ where E is the state energy otherwise reject it
- Go back to step 1 and repeat until done

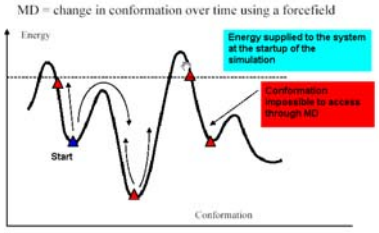
### Monte Carlo Minimization




**Performs a progressive or directed random search**

## Molecular Dynamics (MD)

**In molecular dynamics, energy is supplied to the system, typically using a constant temperature (i.e. constant average kinetic energy).**




### Molecular Dynamics (MD)

- Use Newtonian mechanics to calculate the net force and acceleration experienced by each atom.
- Each atom $i$ is treated as a point with mass $m_i$ and fixed charge $q_i$
- Determine the force $F_i$ on each atom:

$$\vec{F}_i = m_i \frac{d^2\vec{r}_i}{dt^2} = -\vec{\nabla}\nu(\vec{R})$$

- Use positions and accelerations at time $t$ (and positions from $t - \delta t$) to calculate new positions at time $t + \delta t$

Initial velocities ($v_i$)

using the Boltzmann distribution at the given temperature

$$v_i = (m_i/2\pi kT)^{1/2} \exp(-m_i v_i^2/2kT)$$

### Molecular dynamics (MD) simulations

- A deterministic method based on the solution of Newton's equation of motion

$$F_i = m_i a_i$$

for the ith particle; the acceleration at each step is calculated from the negative gradient of the overall potential, using

$$F_i = -\text{grad } V_i = -\nabla V_i$$

**In molecular dynamics forces are derived from a potential energy function *V*, which depend on the particle coordinates:**

$$\mathbf{F}_i = -\nabla V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$$

**The problem of modelling a material can therefore be restated as that of finding a potential function for that material.**

$$V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$$

### Molecular dynamics (MD) simulations

$V_i = \Sigma_k$(energies of interactions between i and all other residues k located within a cutoff distance of $R_c$ from i)

- Derivative of V with respect to the position vector $r_i = (x_i, y_i, z_i)^T$ at each step

$$a_{xi} \sim -\partial V/\partial x_i$$
$$a_{yi} \sim -\partial V/\partial y_i$$
$$a_{zi} \sim -\partial V/\partial z_i$$

***Non-Bonded Interaction Potentials***

- **Electrostatic interactions of the form $E_{ik}(es) = q_i q_k / r_{ik}$**
- **Van der Waals interactions $E_{ij}(vdW) = -a_{ik}/r_{ik}^6 + b_{ik}/r_{ik}^{12}$**

***Bonded Interaction Potentials***

- **Bond stretching $E_i(bs) = (k_{bs}/2)(l_i - l_i^0)^2$**
- **Bond angle distortion $E_i(bad) = (k_\theta/2)(\theta_i - \theta_i^0)^2$**
- **Bond torsional rotation $E_i(tor) = (k_\phi/2) f(\cos\phi_i)$**

**Molecular dynamics (MD) simulations**

### The Verlet algorithm

The most widely used method of integrating the equations of motion is that initially adopted by Verlet [1967] .The method is based on positions r(t), accelerations a (t), and the positions r(t -δt) from the previous step. The equation for advancing the positions reads as

$$r(t+\delta t) = 2r(t)-r(t-\delta t)+ \delta t^2 a(t)$$

The velocities do not appear at all. They have been eliminated by addition of the equations obtained by Taylor expansion about r(t):

$$r(t+\delta t) = r(t) + \delta t\ v(t) + (1/2)\ \delta t^2\ a(t)+ \ldots$$

$$r(t-\delta t) = r(t) - \delta t\ v(t) + (1/2)\ \delta t^2\ a(t)-$$

The velocities are not needed to compute the trajectories, but they are useful for estimating the kinetic energy (and hence the total energy). They may be obtained from the formula

$$v(t)= [r(t+\delta t)-r(t-\delta t)]/2\delta t$$

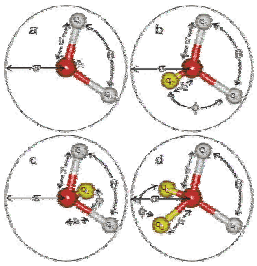
**Trajectory file:** During molecular dynamics (and energy minimization) the coordinates (and velocities) are saved at regular intervals. Such a file is called a trajectory file.

## Water Models

A recent review listed 46 distinct models, so indirectly indicating their lack of success in quantitatively reproducing the properties of real water.

They may, however, offer useful insight into water's behavior.

Models types a, b and c are all planar whereas type d is almost tetrahedral



### Implicit Solvent Models

Water molecules are not included as molecules, but represented by an extra potential on the solvent accessible surface.

·only 50% slower than vacuum calculations

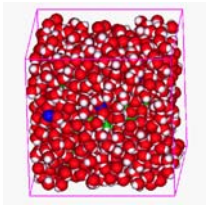
·~10 times faster than explicit water MD

### Explicit Solvent Models

Water molecules are explicitly included as individual molecules.

· Force Fields for water molecules are not trivial …

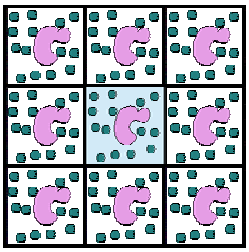
· Computationally expensive …



### Periodic Boundary Conditions (PBC)

· Periodic boundary conditions are used to simulate solvated systems or crystals.

· In solvated systems, PBC prevents that the solvent "evaporates *in silico*"



Building peptides using Z matrices

| | distance | | angle | | dihedral | | connectivity |
|---|---|---|---|---|---|---|---|
| N | 0.0000 | 0 | 0.0000 | 0 | 0.0000 | 0 | 0 0 0 |
| H | 1.0200 | 1 | 0.0000 | 0 | 0.0000 | 0 | 1 0 0 |
| H | 1.0200 | 1 | 104.5368 | 1 | 0.0000 | 0 | 1 2 0 |
| H | 1.0200 | 1 | 104.5368 | 1 | 109.5796 | 1 | 1 2 3 |

0 (end of file)

(1 means optimize, 0 means keep constant, -1 means vary according to a designated pattern)

PDB Representation

```
HETATM  1  C   1    -1.129  1.281 -0.000
HETATM  2  C   2    -2.558  1.772 -0.000
HETATM  3  C   3    -3.519  0.606 -0.000
HETATM  4  H   4    -0.596  1.637  0.890
HETATM  5  H   5    -0.596  1.637 -0.890
HETATM  6  H   6    -2.733  2.392  0.890
HETATM  7  H   7    -2.733  2.392 -0.890
HETATM  8  H   8    -4.558  0.952  0.000
HETATM  9  H   9    -3.359 -0.017 -0.890
HETATM 10  H  10    -3.359 -0.017  0.890
HETATM 11  H  11    -1.110  0.183 -0.000        continued...
```

(blue indicates data columns not utilized/recognized by all software)

PDB Representation contd.

```
CONECT  1  2  4  5 11
CONECT  2  1  3  6  7
CONECT  3  2  8  9 10
CONECT  4  1
CONECT  5  1
CONECT  6  2
CONECT  7  2
CONECT  8  3
CONECT  9  3
CONECT 10  3
CONECT 11  1
END
```

**Atom types (AMBER)**

Table 1. List of Atom Types[a]

| atom | type | description |
|---|---|---|
| carbon | CT | any $sp^3$ carbon |
| | C | any carbonyl $sp^2$ carbon |
| | CA | any aromatic $sp^2$ carbon and (Cz of Arg) |
| | CM | any $sp^2$ carbon, double bonded |
| | CC | $sp^2$ aromatic in 5-membered ring with one substituent + next to nitrogen (Cy in His) |
| | CV | $sp^2$ aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. C6 in His (δ)) |
| | CW | $sp^2$ aromatic in 5-membered ring next to carbon and NH (e.g. C6 in His (ε) and in Trp) |
| | CR | $sp^2$ aromatic in 5-membered ring next to two nitrogens (Cy and Cε in His) |
| | CB | $sp^2$ aromatic at junction of 5- and 6-membered rings (C8 in Trp) and both junction atoms in Ade and Gua |
| | C* | $sp^2$ aromatic in 5-membered ring next to two carbons (e.g. Cγ in Trp) |
| | CN | $sp^2$ junction between 5- and 6-membered rings and bonded to CH and NH (Cε in Trp) |
| | CK | $sp^2$ carbon in 5-membered aromatic between N and N-R (C8 in purines) |
| | CQ | $sp^2$ carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines) |
| nitrogen | N | $sp^2$ nitrogen in amides |
| | NA | $sp^2$ nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp) |
| | NB | $sp^2$ nitrogen in 5-membered ring with lone pair (e.g. N7 in purines) |
| | NC | $sp^2$ nitrogen in 6-membered ring with lone pair (e.g. N3 in purines) |
| | N* | $sp^2$ nitrogen in 5-membered ring with carbon substituent (in purine nucleosides) |
| | N2 | $sp^2$ nitrogen of aromatic amines and guanidinium ions |
| | N3 | $sp^2$ nitrogen |
| oxygen | OW | $sp^3$ oxygen in TIP3P water |
| | OH | $sp^3$ oxygen in alcohols, tyrosine, and protonated carboxylic acids |
| | OS | $sp^3$ oxygen in ethers |
| | O | $sp^2$ oxygen in amides |
| | O2 | $sp^2$ oxygen in anionic acids |
| sulfur | S | sulfur in methionine and cysteine |
| | SH | sulfur in cysteine |
| phosphorus | P | phosphorus in phosphates |
| hydrogen | H | H attached to N |
| | HW | H in TIP3P water |
| | HO | H in alcohols and acids |
| | HS | H attached to sulfur |
| | HA | H attached to aromatic carbon |
| | HC | H attached to aliphatic carbon with no electron-withdrawing substituents |
| | H1 | H attached to aliphatic carbon with one electron-withdrawing substituent |
| | H2 | H attached to aliphatic carbon with two electron-withdrawing substituents |
| | H3 | H attached to aliphatic carbon with three electron-withdrawing substituents |
| | HP | H attached to aromatic carbon ortho to a positively charged group |
| | H4 | H attached to aromatic carbon with one electronegative neighbor (e.g. hydrogen on C5 of Trp, C6 of Tby) |
| | H5 | H attached to aromatic carbon with two electronegative neighbors (e.g. H8 of Ade and Gua and H2 of Ade) |

**Bond Parameters**

Bond Parameters

| bond | $K_r$[b] | $r_{eq}$[c] | bond | $K_r$ | $r_{eq}$ | bond | $K_r$ | $r_{eq}$ | bond | $K_r$ | $r_{eq}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C–CA | 469.0 | 1.409 | CA–HA | 367.0 | 1.080 | CM–HA | 367.0 | 1.080 | CT–S | 227.0 | 1.810 |
| C–CB | 447.0 | 1.419 | CA–N2 | 481.0 | 1.340 | CM–N* | 448.0 | 1.365 | CT–SH | 237.0 | 1.810 |
| C–CM | 410.0 | 1.444 | CA–NA | 427.0 | 1.381 | CN–NA | 428.0 | 1.380 | CV–H4 | 367.0 | 1.080 |
| C–CT | 317.0 | 1.522 | CA–NC | 483.0 | 1.339 | CQ–H5 | 367.0 | 1.080 | CV–NB | 410.0 | 1.394 |
| C–N | 490.0 | 1.335 | CB–CB | 520.0 | 1.370 | CQ–NC | 502.0 | 1.324 | CW–H4 | 367.0 | 1.080 |
| C–N* | 424.0 | 1.383 | CB–CN | 447.0 | 1.419 | CR–H5 | 367.0 | 1.080 | CW–NA | 427.0 | 1.381 |
| C–NA | 418.0 | 1.388 | CB–N* | 436.0 | 1.374 | CR–NA | 477.0 | 1.343 | H–N | 434.0 | 1.010 |
| C–NC | 457.0 | 1.358 | CB–NB | 414.0 | 1.391 | CR–NB | 488.0 | 1.335 | H–N* | 434.0 | 1.010 |
| C–O | 570.0 | 1.229 | CB–NC | 461.0 | 1.354 | CT–CT | 310.0 | 1.526 | H–N2 | 434.0 | 1.010 |
| C–O2 | 656.0 | 1.250 | CC–CT | 317.0 | 1.504 | CT–F | 367.0 | 1.380 | H–N3 | 434.0 | 1.010 |
| C–OH | 450.0 | 1.364 | CC–CV | 512.0 | 1.375 | CT–H1 | 340.0 | 1.090 | H–NA | 434.0 | 1.010 |
| C*–CB | 388.0 | 1.459 | CC–CW | 518.0 | 1.371 | CT–H2 | 340.0 | 1.090 | HO–OH | 553.0 | 0.960 |
| C*–CT | 317.0 | 1.495 | CC–NA | 422.0 | 1.385 | CT–H3 | 340.0 | 1.090 | HO–OS | 553.0 | 0.960 |
| C*–CW | 546.0 | 1.352 | CC–NB | 410.0 | 1.394 | CT–HC | 340.0 | 1.090 | HS–SH | 274.0 | 1.336 |
| C*–HC | 367.0 | 1.080 | CC–NC | 457.0 | 1.340 | CT–HP | 340.0 | 1.090 | O2–P | 525.0 | 1.480 |
| CA–CA | 469.0 | 1.400 | CK–N* | 440.0 | 1.371 | CT–N | 337.0 | 1.449 | OH–P | 230.0 | 1.610 |
| CA–CB | 469.0 | 1.404 | CK–NB | 529.0 | 1.304 | CT–N* | 337.0 | 1.475 | OS–P | 230.0 | 1.610 |
| CA–CM | 427.0 | 1.433 | CM–CM | 549.0 | 1.350 | CT–N2 | 337.0 | 1.463 | OW–HW | 553.0 | 0.9572 |
| CA–CN | 469.0 | 1.400 | CM–CT | 317.0 | 1.510 | CT–N3 | 367.0 | 1.471 | S–S | 166.0 | 2.038 |
| CA–CT | 317.0 | 1.510 | CM–H4 | 367.0 | 1.080 | CT–OH | 320.0 | 1.410 | | | |
| CA–H4 | 367.0 | 1.080 | CM–H5 | 367.0 | 1.080 | CT–OS | 320.0 | 1.410 | | | |

**Angle Parameters**

Angle Parameters

| angle | $K_\theta$[d] | $\theta_{eq}$[e] | angle | $K_\theta$ | $\theta_{eq}$ | angle | $K_\theta$ | $\theta_{eq}$ | angle | $K_\theta$ | $\theta_{eq}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C–CA–CA | 63.0 | 120.00 | CA–CT–HC | 50.0 | 109.50 | CN–NA–H | 30.0 | 123.10 | H1–CT–N2 | 50.0 | 109.50 |
| C–CA–HA | 35.0 | 120.00 | CA–N2–CT | 50.0 | 123.20 | CR–NA–CW | 70.0 | 120.00 | H1–CT–OH | 50.0 | 109.50 |
| C–CB–CB | 63.0 | 119.20 | CA–N2–H | 35.0 | 120.00 | CR–NA–H | 30.0 | 120.00 | H1–CT–OS | 50.0 | 109.50 |
| C–CB–NB | 70.0 | 130.00 | CA–NA–H | 30.0 | 118.00 | CR–NB–CV | 70.0 | 117.00 | H1–CT–S | 50.0 | 109.50 |
| C–CM–CM | 63.0 | 120.70 | CA–NC–CB | 70.0 | 112.20 | CT–C–N | 70.0 | 116.60 | H1–CT–SH | 50.0 | 109.50 |
| C–CM–CT | 70.0 | 119.70 | CA–NC–CQ | 70.0 | 118.60 | CT–S–O | 80.0 | 120.40 | H2–CT–H2 | 35.0 | 109.50 |
| C–CM–H4 | 35.0 | 119.70 | CB–C–NA | 70.0 | 111.30 | CT–C–O2 | 70.0 | 117.00 | H2–CT–N* | 50.0 | 109.50 |
| C–CM–HA | 35.0 | 119.70 | CB–C–O | 80.0 | 128.80 | CT–C–CW | 70.0 | 125.00 | H2–CT–OS | 50.0 | 109.50 |
| C–CT–CT | 63.0 | 111.10 | CB–C*–CT | 70.0 | 128.60 | CT–CC–CV | 70.0 | 120.00 | H4–CM–N* | 35.0 | 119.10 |
| C–CT–H1 | 50.0 | 109.50 | CB–C*–CW | 63.0 | 106.40 | CT–CC–CW | 70.0 | 120.00 | H4–CV–NB | 35.0 | 120.00 |
| C–CT–HC | 50.0 | 109.50 | CB–CA–H4 | 35.0 | 120.00 | CT–CC–NA | 70.0 | 120.00 | H4–CW–NA | 35.0 | 120.00 |
| C–CT–HP | 50.0 | 109.50 | CB–CA–HA | 35.0 | 120.00 | CT–CC–NB | 70.0 | 120.00 | H5–CK–N* | 35.0 | 123.05 |
| C–CT–N | 63.0 | 110.10 | CB–CN–N2 | 70.0 | 123.50 | CT–CT–CT | 40.0 | 109.50 | H5–CR–NB | 35.0 | 123.05 |
| C–CT–N3 | 80.0 | 111.20 | CB–CA–NC | 70.0 | 117.30 | CT–CT–H1 | 50.0 | 109.50 | H5–CQ–NC | 35.0 | 115.45 |
| C–N–CT | 50.0 | 121.90 | CB–CB–N* | 70.0 | 106.20 | CT–CT–H2 | 50.0 | 109.50 | H5–CR–NA | 35.0 | 120.00 |
| C–N–H | 30.0 | 120.00 | CB–CB–NB | 70.0 | 110.40 | CT–CT–HC | 50.0 | 109.50 | H5–CR–NB | 35.0 | 120.00 |
| C–N*–CM | 70.0 | 121.60 | CB–CB–NC | 70.0 | 127.70 | CT–CT–HP | 50.0 | 109.50 | HC–CT–HC | 35.0 | 109.50 |
| C–N*–CT | 70.0 | 117.60 | CB–CN–NA | 70.0 | 104.40 | CT–CT–N | 80.0 | 109.70 | HO–OH–P | 45.0 | 108.50 |
| C–N*–H | 30.0 | 119.20 | CB–N*–CK | 70.0 | 105.40 | CT–CT–N* | 50.0 | 109.50 | HP–CT–HP | 35.0 | 109.50 |
| C–NA–C | 70.0 | 126.40 | CB–N*–CT | 70.0 | 125.80 | CT–CT–N2 | 80.0 | 111.20 | HP–CT–N3 | 50.0 | 109.50 |
| C–NA–CA | 70.0 | 125.20 | CB–N*–H | 30.0 | 125.80 | CT–CT–N3 | 80.0 | 111.20 | H5–SH–HS | 35.0 | 92.07 |
| C–NA–H | 30.0 | 116.80 | CB–NB–CK | 70.0 | 103.80 | CT–CT–OH | 50.0 | 109.50 | HW–OW–HW | 100.0 | 104.52 |
| C–NC–CA | 70.0 | 120.50 | CB–NB–CQ | 70.0 | 111.00 | CT–CT–OS | 50.0 | 109.50 | N–C–O | 80.0 | 122.90 |
| C–OH–HO | 35.0 | 113.00 | CC–CT–CT | 63.0 | 113.10 | CT–CT–S | 50.0 | 114.70 | N*–C–NA | 70.0 | 115.40 |
| C*–CB–CA | 63.0 | 134.90 | CC–CT–HC | 50.0 | 109.50 | CT–CT–SH | 50.0 | 108.60 | N*–C–NC | 70.0 | 118.00 |
| C*–CB–CN | 63.0 | 108.80 | CC–CV–H4 | 35.0 | 120.00 | CT–N–CT | 50.0 | 118.00 | N*–C–O | 80.0 | 120.90 |
| C*–CT–CT | 63.0 | 115.60 | CC–CV–NB | 70.0 | 120.00 | CT–N–H | 30.0 | 118.04 | N*–CB–NC | 70.0 | 126.20 |
| C*–CT–HC | 50.0 | 109.50 | CC–CW–H4 | 35.0 | 120.00 | CT–N2–H | 35.0 | 118.40 | N*–CK–NB | 70.0 | 113.90 |
| C*–CW–H4 | 35.0 | 120.00 | CC–CW–NA | 70.0 | 120.00 | CT–N3–H | 50.0 | 109.50 | N*–CT–OS | 50.0 | 109.50 |
| C*–CW–NA | 70.0 | 108.70 | CC–NA–CR | 70.0 | 120.00 | CT–OH–HO | 55.0 | 108.50 | N2–CA–N2 | 70.0 | 120.00 |
| CA–C–CA | 63.0 | 120.00 | CC–NA–H | 30.0 | 120.00 | CT–OS–CT | 60.0 | 109.50 | N2–CA–NA | 70.0 | 114.00 |
| CA–C–OH | 70.0 | 120.00 | CC–NB–CR | 70.0 | 117.00 | CT–OS–P | 100.0 | 120.50 | N2–CA–NC | 70.0 | 119.50 |
| CA–CA–CA | 63.0 | 120.00 | CK–N*–CT | 70.0 | 128.80 | CT–S–CT | 62.0 | 98.90 | NA–C–O | 80.0 | 120.80 |
| CA–CA–CB | 63.0 | 120.00 | CK–N*–H | 30.0 | 128.80 | CT–S–S | 68.0 | 103.70 | NA–CA–NC | 70.0 | 123.30 |
| CA–CA–CN | 63.0 | 120.00 | CM–C–NA | 70.0 | 114.10 | CT–SH–HS | 43.0 | 96.00 | NA–CR–NA | 70.0 | 120.00 |
| CA–CA–CT | 70.0 | 120.00 | CM–C–O | 80.0 | 125.30 | CV–CC–NA | 70.0 | 120.00 | NA–CR–NB | 70.0 | 120.00 |
| CA–CA–H4 | 35.0 | 120.00 | CM–CA–N2 | 70.0 | 120.10 | CW–CC–NA | 70.0 | 120.00 | NC–C–O | 80.0 | 122.80 |

**Torsion Parameters**

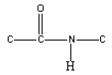
Torsional Parameters

| torsion | no. of paths[f] | $V_n/2$[g] | $\gamma$[h] | $n$[i] | torsion | no. of paths[f] | $V_n/2$[g] | $\gamma$[h] | $n$[i] |
|---|---|---|---|---|---|---|---|---|---|
| X–C–CA–X | 4 | 14.50 | 180.0 | 2.0 | X–CT–OH–X | 3 | 0.50 | 0.0 | 3.0 |
| X–C–CB–X | 4 | 12.00 | 180.0 | 2.0 | X–CT–OS–X | 3 | 1.15 | 0.0 | 3.0 |
| X–C–CM–X | 4 | 8.70 | 80.0 | 2.0 | X–CT–S–X | 3 | 1.00 | 0.0 | 3.0 |
| X–C–CT–X | 4 | 0.00 | 0.0 | 2.0 | X–CT–SH–X | 3 | 0.75 | 0.0 | 3.0 |
| X–C–N–X | 4 | 10.00 | 180.0 | 2.0 | X–CV–NB–X | 2 | 4.80 | 180.0 | 2.0 |
| X–C–N*–X | 4 | 5.80 | 180.0 | 2.0 | X–CW–NA–X | 4 | 6.00 | 180.0 | 2.0 |
| X–C–NA–X | 4 | 5.40 | 180.0 | 2.0 | X–OH–P–X | 3 | 0.75 | 0.0 | 3.0 |
| X–C–NC–X | 2 | 8.00 | 180.0 | 2.0 | X–OS–P–X | 3 | 0.75 | 0.0 | 3.0 |
| X–C–OH–X | 2 | 1.80 | 180.0 | 2.0 | C–N–CT–C | 1 | 0.0 | 0.0 | +4.0 |
| X–C*–CB–X | 4 | 6.70 | 180.0 | 2.0 | C–N–CT–C | 1 | 0.0 | 180.0 | −3.0 |
| X–C*–CT–X | 6 | 0.00 | 0.0 | 2.0 | C–N–CT–C | 1 | 0.20 | 180.0 | −2.0 |
| X–C*–CW–X | 4 | 26.10 | 180.0 | 2.0 | C–N–CT–C | 1 | 0.00 | 180.0 | 1.0 |
| X–CA–CA–X | 4 | 14.50 | 180.0 | 2.0 | CT–CT–C–N | 1 | 0.100 | 0.0 | +4.0 |
| X–CA–CB–X | 4 | 14.00 | 180.0 | 2.0 | CT–CT–C–N | 1 | 0.000 | 0.0 | −3.0 |
| X–CA–CM–X | 4 | 10.20 | 180.0 | 2.0 | CT–CT–C–N | 1 | 0.07 | 0.0 | −2.0 |
| X–CA–CN–X | 4 | 14.50 | 180.0 | 2.0 | CT–CT–C–N | 1 | 0.000 | 180.0 | 1.0 |
| X–CA–CT–X | 6 | 0.00 | 0.0 | 2.0 | CT–CT–N–C | 1 | 0.50 | 180.0 | +4.0 |
| X–CA–N2–X | 4 | 9.60 | 180.0 | 2.0 | CT–CT–N–C | 1 | 0.15 | 180.0 | −3.0 |
| X–CA–NA–X | 4 | 8.00 | 180.0 | 2.0 | CT–CT–N–C | 1 | 0.00 | 180.0 | −2.0 |
| X–CA–NC–X | 2 | 9.60 | 180.0 | 2.0 | CT–CT–N–C | 1 | 0.53 | 0.0 | 1.0 |
| X–CB–CB–X | 4 | 21.80 | 180.0 | 2.0 | CT–CT–OS–CT | 1 | 0.383 | 0.0 | +3.0 |
| X–CB–CN–X | 4 | 12.00 | 180.0 | 2.0 | CT–CT–OS–CT | 1 | 0.1 | 180.0 | 2.0 |
| X–CB–N*–X | 4 | 6.60 | 180.0 | 2.0 | CT–S–S–CT | 1 | 0.60 | 0.0 | 3.0 |
| X–CB–NB–X | 2 | 5.10 | 180.0 | 2.0 | CT–S–S–CT | 1 | 3.50 | 0.0 | −2.0 |
| X–CC–CT–X | 6 | 8.30 | 180.0 | 2.0 | H–N–C–O | 1 | 2.50 | 180.0 | +2.0 |
| X–CC–CV–X | 4 | 0.00 | 0.0 | 2.0 | H–N–C–O | 1 | 2.00 | 0.0 | 1.0 |
| X–CC–CW–X | 4 | 20.00 | 180.0 | 2.0 | N–CT–C–N | 1 | 0.40 | 180.0 | +4.0 |
| X–CC–NA–X | 4 | 21.50 | 180.0 | 2.0 | N–CT–C–N | 1 | 0.0 | 0.0 | −3.0 |
| X–CC–NA–X | 4 | 5.80 | 180.0 | 2.0 | N–CT–C–N | 1 | 1.35 | 180.0 | −2.0 |
| X–CC–NB–X | 2 | 4.80 | 180.0 | 2.0 | N–CT–C–N | 1 | 0.75 | 180.0 | 1.0 |
| X–CK–N*–X | 4 | 6.80 | 180.0 | 2.0 | OH–CT–CT–OH | 1 | 0.144 | 0.0 | −3.0 |
| X–CK–NB–X | 2 | 20.00 | 180.0 | 2.0 | OH–CT–CT–OH | 1 | 1.00 | 0.0 | 2.0 |
| X–CM–CM–X | 4 | 26.60 | 180.0 | 2.0 | OH–P–OS–CT | 1 | 0.25 | 0.0 | −3.0 |
| X–CM–CT–X | 6 | 0.00 | 0.0 | 3.0 | OH–P–OS–CT | 1 | 1.20 | 0.0 | 2.0 |
| X–CM–N*–X | 4 | 7.40 | 180.0 | 2.0 | OS–CT–CT–OH | 1 | 0.144 | 0.0 | −3.0 |
| X–CN–NA–X | 4 | 6.10 | 180.0 | 2.0 | OS–CT–CT–OH | 1 | 1.00 | 0.0 | 2.0 |
| X–CO–NC–X | 2 | 13.60 | 180.0 | 2.0 | OS–CT–CT–OH | 1 | 0.144 | 0.0 | −3.0 |

## Improper Torsions

Improper Torsions

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |



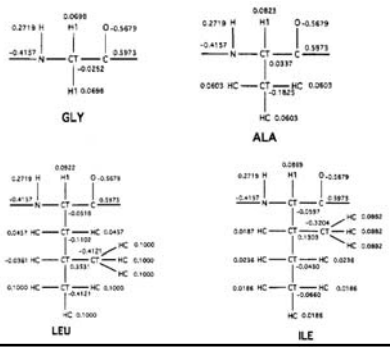
## Van der Waals (LJ) Parameters

Van der Waals Parameters

| atom type | $R^{vdw}$ | $\epsilon^{k}$ | atom type | $R^{vdw}$ | $\epsilon^{k}$ | atom type | $R^{vdw}$ | $\epsilon^{k}$ | atom type | $R^{vdw}$ | $\epsilon^{k}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 1.9080 | 0.0860 | H2 | 1.2870 | 0.0157 | HS | 0.6000 | 0.0157 | O2 | 1.6612 | 0.2100 |
| CA | 1.9080 | 0.0860 | H3 | 1.1870 | 0.0157 | HW | 0.0000 | 0.0000 | OH | 1.7210 | 0.2104 |
| CM | 1.9080 | 0.0860 | H4 | 1.4090 | 0.0150 | IP | 1.8680 | 0.00277 | OS | 1.6837 | 0.1700 |
| $C_3$ | 3.3950 | 0.0000806 | H5 | 1.3590 | 0.0150 | K | 2.6580 | 0.000328 | OW | 1.7683 | 0.1520 |
| CT | 1.9080 | 0.1094 | HA | 1.4590 | 0.0150 | Li | 1.1370 | 0.0183 | P | 2.1000 | 0.2000 |
| F | 1.75 | 0.061 | HC | 1.4870 | 0.0157 | $N^*$ | 1.8240 | 0.1700 | Rb | 2.9560 | 0.000017 |
| H | 0.6000 | 0.0157 | HO | 0.0000 | 0.0000 | $N3^*$ | 1.875 | 0.1700 | S | 2.0000 | 0.2500 |
| H1 | 1.3870 | 0.0157 | HP | 1.1000 | 0.0157 | O | 1.6612 | 0.2100 | SH | 2.0000 | 0.2500 |

$$V_{LJ} = \sum_{i<j} 4\varepsilon_{ij}\left[\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right]$$

## Atomic Partial Charges




## Typical Time Scales ....

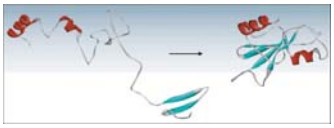
- Bond stretching: $10^{-14}$ - $10^{-13}$ sec.
- Elastic vibrations: $10^{-12}$ - $10^{-11}$ sec.
- Rotations of surface sidechains: $10^{-11}$ - $10^{-10}$ sec.
- Hinge bending: $10^{-11}$ - $10^{-7}$ sec.
- Rotation of buried side chains: $10^{-4}$ - 1 sec.
- Protein folding: $10^{-6}$ - $10^{2}$ sec.

Timescale in MD:

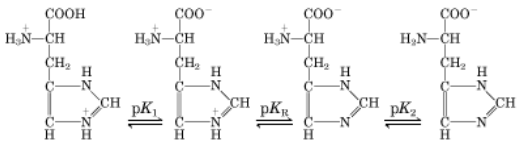
- A Typical timestep in MD is 1 fs ($10^{-15}$ sec)

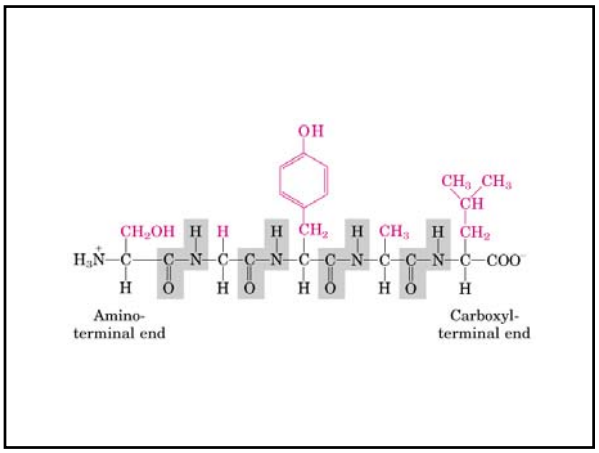
(ideally 1/10 of the highest frequency vibration)

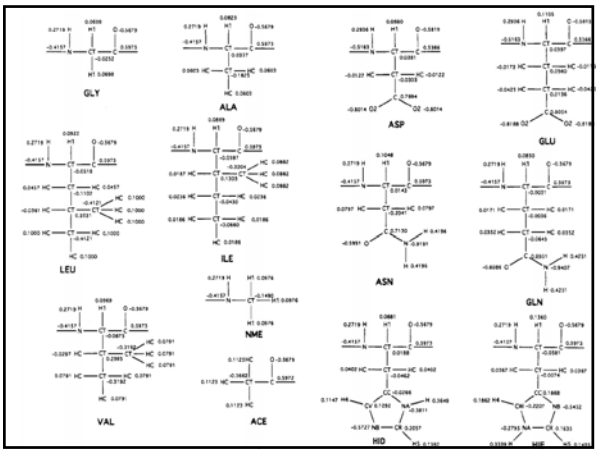
## *Ab initio* protein folding simulation

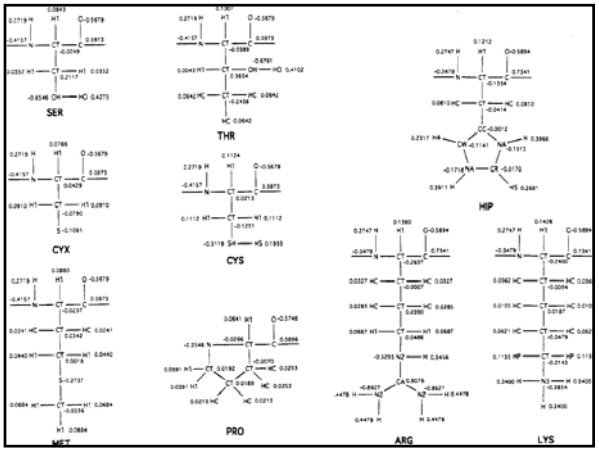


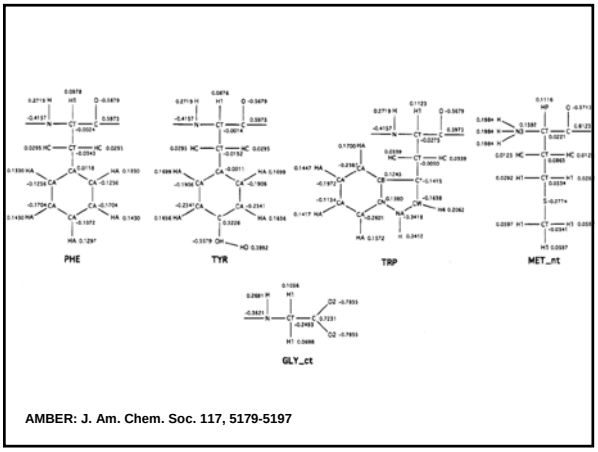
| | |
|---|---|
| Physical time for simulation | $10^{-4}$ seconds |
| Typical time-step size | $10^{-15}$ seconds |
| Number of MD time steps | $10^{11}$ |
| Atoms in a typical protein and water simulation | 32,000 |
| Approximate number of interactions in force calculation | $10^{9}$ |
| Machine instructions per force calculation | 1000 |
| Total number of machine instructions | $10^{23}$ |
| BlueGene capacity (floating point operations per second) | 1 pentaflop ($10^{15}$) |

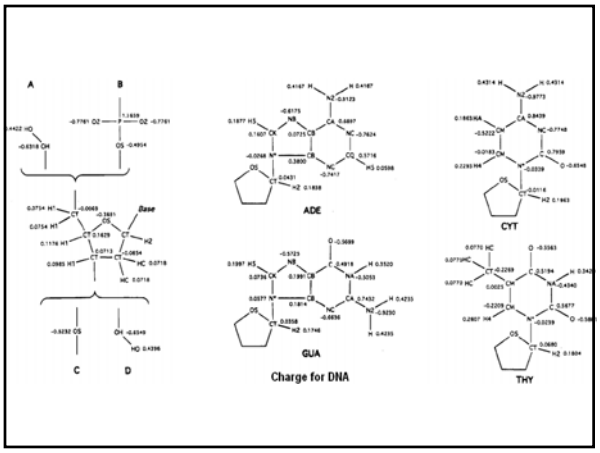


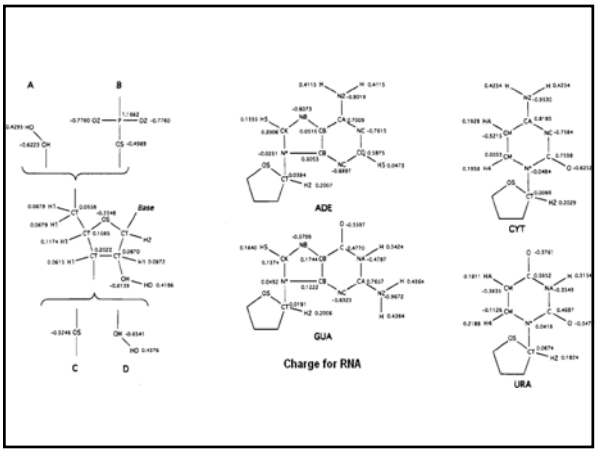

Amino-
terminal end
Carboxyl-
terminal end


GLY
ALA
ASP
GLU
LEU
ILE
ASN
GLN
VAL
NME
ACE
HID


SER
THR
HIP
CYX
CYS
MET
PRO
ARG
LYS


PHE
TYR
TRP
MET_nt
GLY_ct
AMBER: J. Am. Chem. Soc. 117, 5179-5197


Base
ADE
GUA
CYT
THY
Charge for DNA


Base
ADE
GUA
CYT
URA
Charge for RNA