Lecture 5: Molecular Mechanics Force Fields

I. Introduction

Molecular simulations of molecules with bonded atoms require three separate types of information to complete the molecular model:

1. Atomic coordinates and names.
2. The covalent connectivity; which atoms are bonded to which others.
3. Potential Energy function.

These three items together constitute a complete molecular mechanics model and number three is the **force field** (FF) used to calculate interaction energies during simulation. There are a number of "complete" empirical force fields in common use today. The have acronymous names like AMBER, CHARMM, CEDAR, GROMOS, and OPLS and they share similar concepts and analytical functions. Although the important distinctions between these force fields are the parameters in the potential energy function, each force field also has its own version of atom names. We will use the CHARMM force field in the lab simulations of a protein in water. However, the water model in CHARMM is derived from the OPLS water model.

One of the take-home messages today is that force fields are useful but imperfect. We will come to appreciate the many compromises that are made in order to obtain a useful simulation. If we waited until the perfect force field was available, little progress would be made. Although we will use a complete force field in the lab and make some conclusions about the results, it is important to know the limitations of force fields.

1. Initial Positions.

Most simulation programs use a variation of the PDB format for listing the names and x, y, z coordinates of atoms. Beware that there may be slight formatting differences between programs and some programs use completely different formats. A common headache for computational structural biologists is the constant inter-conversion of file formats! If you do research in this field you need a collection of utilities (awk, python, etc.) to convert back and forth.

In addition, the atomic naming is specific to each force field. In a PDB file each atom in an amino acid residue has a different name. In force field descriptions, all atoms of the same *chemical type* have the same name. This makes comparison of the PDB file atoms and molecular mechanics model atoms confusing.

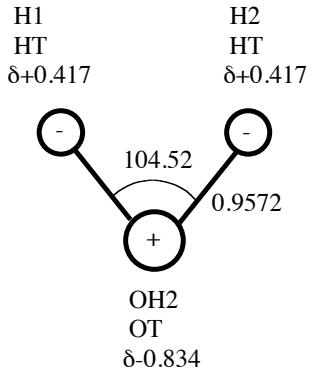
2. Covalent connectivity.

An original idea for empirical force fields was that a set of "**transferable intermolecular potential functions**" (TIPS) for common chemical groups could be described and calibrated and then used to build potential functions for more complicated models. Molecular mechanics models of water, methanol, ethanol, simple carboxylic acids and amides were originally devised. In the early 1980's William Jorgensen came out with a

series of TIPS which were successful in simulating structural and energetic properties of water and small organic liquids in good agreement with experimental data. This work grew into the force field OPLS (Optimized Parameters for Liquid Simulation). We will look at the TIPS model of water as an example (and it is still used in the CHARMM FF).

The connectivity (bonding) of a molecular mechanics (MM) model is usually described in the "**topology**" section of a force field list. Below is the model for the TIPS3 water as a figure and as a text extract from the CHARMM force field topology for this molecule. The "3" in the name derives from the fact that there are three sites of partial charge as shown below; one on each atom. The TIPS3 (or TIP3) water model is one of the most widely used water models. There are several MM models for water which have slightly different bond lengths, angles and charge configurations. Notice that the topology information provides mass, connectivity and partial charge; the bond lengths and bond angles are provided in the "**parameter**" section which we will get to below.

Notice also the atom nomenclature. Here the first atom names (OH2, H1, H2) are PDB names; the second names (OT, HT) are specific for this force field. This renaming was done so that atoms with similar bonding and charge could be grouped together.




```
MASS     4 HT     1.00800 H ! TIPS3P WATER HYDROGEN
MASS    75 OT    15.99940 O ! TIPS3P WATER OXYGEN

RESI TIP3          0.000
GROUP
ATOM OH2  OT     -0.834
ATOM H1   HT      0.417
ATOM H2   HT      0.417
BOND OH2 H1 OH2 H2
ANGLE H1 OH2 H2
```

The GROUP section defines a group of atoms that have a net zero charge. Here the two partially positive hydrogens balance the partially negative oxygen.

The topology section indicates that oxygen (OH2) is bonded to both H1 and H2. The BOND lengths and bond ANGLE values will be listed in the parameter file described below.

For comparison, the topology of the amino acid residue alanine is shown below. Again, a GROUP is defined as a collection of partial charges that sum to zero. Notice that the

three methyl hydrogens have different PDB names (HB1, HB2, HB3) because they each have different coordinates but the same MM model name (HA) because they all have the same charge and mass. The last three lines in the topology section describe the atoms that are bonded to each other and the type of bond between them (single or double).

```
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
BOND CB CA  N  HN  N  CA
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
DOUBLE O  C
```

To summarize: The **topology** of a molecule describes its mass, bonding connectivity and charge. The energetic parameters associated with each atom and bond are defined in the **parameter** section of a force field that is described next. (Note: Charge is also an energy parameter but traditionally is included in the topology sections – don't ask why).

3. Energy Functions.

The immediate purpose of molecular simulations is to generate a Boltzmann distribution of energy states. (!) Therefore an understanding of how energy is calculated is important for understanding molecular simulations. We have previously discussed the Lennard-Jones potential as a force field for the Argon simulations, and we learned in the LINUS simulation that a force field could be a simple scoring function accounting for the presence of hydrogen bonds or a pseudo function mimicking the hydrophobic effect. Before we begin a discussion of a "complete" empirical force field it would be good to review some terminology.

> **FREE ENERGY AND POTENTIAL ENERGY**. At equilibrium, the free energy of an ensemble is at a minimum. In the NVT ensemble where the volume is constant and no pressure work can be done the Helmholtz free energy (A) is
>
> $$A = E - TS$$
> $$A = -kT\ln Z$$
>
> where E is the average total energy (kinetic and potential), T is temperature, S is entropy and Z is the partition function. In the NPT ensemble, the Gibbs free energy (G) is
>
> $$G = E + PV - TS$$
> $$G = -kT\ln Z$$

where P is pressure, V is volume and $E + PV = H$, the enthalpy. H is equivalent to E in most biochemical systems where $\Delta V = 0$ and both represent the average energy of the system. Calculation of **free energy** requires a complete enumeration of the possible states in the ensemble, i.e. it requires a calculation of the partition function ($G=-kT\ln Z$). In molecular mechanics models the empirical energy function describes only the **potential energy** of the system. Entropy is not explicitly considered in the calculation of energy during a simulation; rather, entropy is inherent in the number of states found in a complete ensemble. Special averaging methods over many states of the ensemble makes it possible get changes in free energy (ΔG or ΔA) from molecular simulations using the potential energy force fields discussed today, but we will not include this method in the course.

A force field that includes empirical terms to account for entropy will be introduced later. This latter force field attempts to estimate the **free energy** of a single system directly. Today, we focus only on force fields that attempt to estimate the potential energy of a system at every step.

An empirical potential energy function commonly used in molecular mechanics models is composed of different terms to account for energies associated with the following:

- Bond lengths
- Bond angles
- Bond torsion angles
- van der Waals interactions
- Electrostatic interactions

In equation form the potential function is

$$E(\mathbf{r}) = \sum_{bond} K_b(b - b_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{torsion} K_\Phi [1 + \cos(n(\Phi - \Phi_0)] +$$

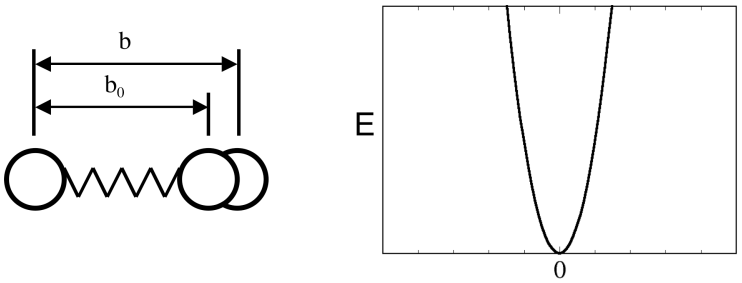
$$\sum_{nonbonded} \left(\varepsilon_{ij}[(R_{min} / R_{ij})^{12} - 2(R_{min} / R_{ij})^6] + q_i q_j / 4\pi D r_{ij}\right)$$

where the individual symbols are explained below. This function is hard wired in the code of the simulation program; the variables are assigned values from the **parameter** list that is input to the code as a parameter file.

a. Bond length – Springs.

$$E = K_b(b - b_0)^2$$

where K_b = spring constant (~ 500 kcal/mol/Å^2) (large!) and b, b_0 are depicted in the figure below; b_0 is the equilibrium length of the bond, b is the instantaneous bond length. The large spring constant provides a narrow harmonic energy well. Bonds may *not* be stretched very much.



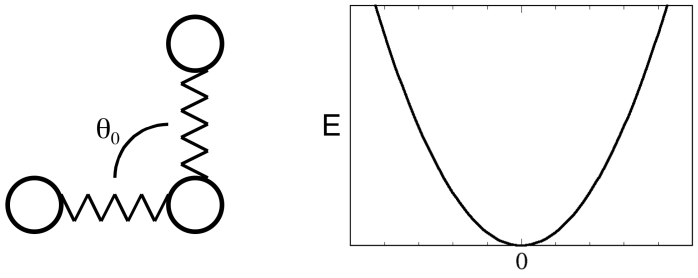
The relevant extract from a CHARMM parameter file is shown below. The parameters given are K_b and b_o.

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
OT   HT    450.000     0.9572
```

b. Bond Angles

$$E = K_\theta (\theta - \theta_0)^2$$

where $K_\theta \sim 50$ kcal/mol/Å^2 and θ is defined in the figure below. A smaller spring constant provides a shallower energy well and allows angles to vary proportionately more than bond lengths.



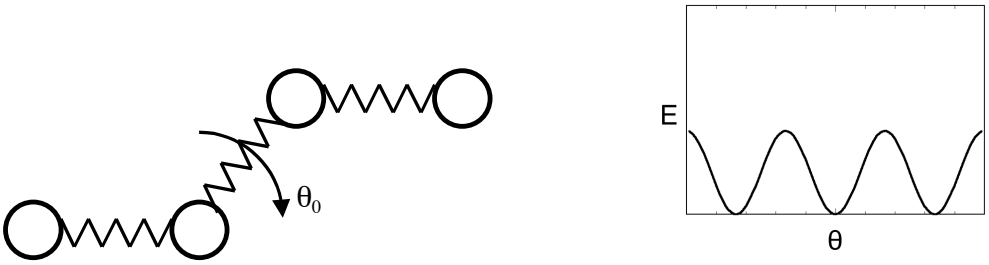
The relevant extract from a CHARMM parameter file is shown below. The parameters given are K_θ and θ_0 .

```
ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!
!atom types     Ktheta    Theta0
HT   OT   HT     55.000   104.5200
```

c. Dihedral angles (torsion angles)

$$E = K_\chi[1 + (\cos(n\theta - \delta))]$$

where $K_\chi \sim 1$, *n* is the symmetry of the rotor (e.g. 3 for methyl groups) and δ is the phase. Notice in the plot that, initially, all wells and barriers are equal.



Multiple dihedrals are calculated for each bond to account for the different groups on the central two atoms above and summed. The result is different well depths and barrier heights in the above plot (*cf*. the energy plot for butane that we have used previously). The following table may help to understand how the phase accounts for different bonding to the central two atoms. The phase of each four-atom group is given in the parameter file.

Dihedral Parameters			
Multiplicity, *n*	Phase, δ	Location of minima	Notes
1	0	180	yields trans
1	180	0	yields cis
2	0	90, 270	
2	180	0, 180	enforces planarity
3	0	60, 180, 300	staggered sp3 carbons
3	180	0, 120, 240	eclipsed sp3 carbons

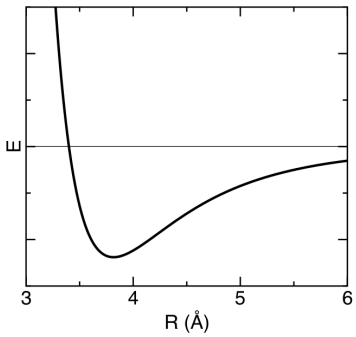
The relevant extract from a CHARMM parameter file is shown below. The parameters given are K_χ, and n and δ.

```
DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi    n   delta
C    CT1  NH1  C        0.2000  1   180.00
CC   CT1  CT2  CA       0.0400  3     0.00
```

d. Lennard-Jones term.

$$E = \varepsilon_{ij}[(R_{min} / R_{ij})^{12} - 2(R_{min} / R_{ij})^{6}]$$

Where ε_{ij} is the energy of interaction between two atoms at the most favorable distance R_{min}. This is a different form of the L-J expression that uses R_{min} instead of σ; it is equivalent to the form of the expression we learned earlier. Remember from your homework assignment that the LJ interaction energy is on the order of only 0.2 kcal/mol per atom pair and decays quickly because of the $1/r^6$ relationship.



The parameter file (see below) gives the ε (epsilon) and $R_{min}/2$ (VDW radius) for the interaction of each atom type with itself. The extract from the parameter file below shows how the averaging is done for *two different atoms*. The combination $R_{min,i,j}$ is simply an **arithmetic average**. The average of two different epsilon values, $Eps_{i,j}$, is obtained by **geometric averaging**: sqrt(eps_i * eps_j).

```
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom             epsilon      Rmin/2
HT               -0.046000     0.224500
OT               -0.152100     1.768200
```

The use of the L-J potential assumes that the sum of pairwise interactions captures the overall multibody interaction of an atom with all of its neighbors. That is, we assume that the pairwise interaction is not influenced by the type and number of neighboring atoms. Clearly this assumption cannot be true. Yet, to go beyond the two body pairwise interaction scheme would be prohibitively expensive in computer time. The assumption seems to work because the distance dependence of van der Waals interactions falls off so quickly ($1/r^6$). (What two factors fall off with a $1/r^3$ dependence and sum to give the $1/r^6$ total dependence?)

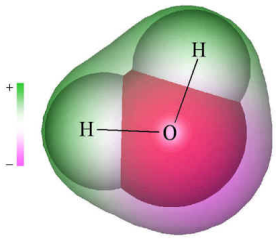
Values of R_{min} and epsilon for a particular atom type are not consistent between force fields. The table below collects some values of $R_{min}/2$ used in a variety of studies. The values in columns A-F were derived from statistical analysis of atomic distances in the PDB or calculated from the densities and volumes of liquids. The distances represent the average distance between atoms of the respective type (divided by 2) and they should therefore represent the van der Waals radius, R_{min}. (This analysis assumes that structures in the PDB represent a Boltzmann distribution of atom to atom distances, i.e. an equilibrium set of distances.) The values in columns G & H are from two MM force fields. Notice that the radii used in molecular mechanics models are larger than apparent radii found in liquids and crystal structures. This appears to be necessary to mimic the

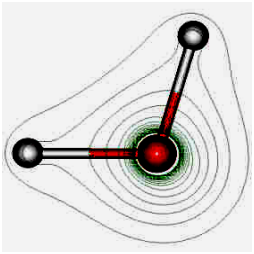
multibody effects when performing a simulation. Remember that the non-bonding interactions are summed over all neighboring atoms and pairwise interactions may not be the best way to capture how neighbors affect these interactions.

Symbol	Type	Designation	Liquid & Xtal Database							Molec. Mech.		
			A	B	C	D	E	F	Ave.	G	H	Ave.
H	Average		1.2	1.17	1.2	1.10	1.17		**1.17**			
-H	Polar	HN								0.00	0.22	
-H	Nonpolar	HB*,HD1,HA								0.00	1.32	
C	Average		1.85	1.80	1.91	1.77	1.75	1.85	**1.82**	2.14	2.11	**2.13**
>C=	Carbonyl	C						1.75		2.10	2.00	
-CH=	Aromatic	CG,CD*,CE*,CZ	1.7		1.74			1.82		2.10	1.99	
>CH-	Aliph sp3	CA						1.90		2.13	2.28	
-CH2-	Aliph sp3	CB			2.00			1.90		2.19	2.17	
-CH3	Aliph sp3	CG*,CD*,CE,CM*	2.0		2.00			1.92		2.19	2.17	
N	Average		1.5	1.58	1.60	1.64	1.55	1.66	**1.59**	1.82	1.85	**1.84**
>NH	Peptide	N						1.70		1.82	1.85	
-NH2	Amide	ND2,NE2						1.62		1.82	1.85	
-NH3+	Amonium	NZ						1.67		1.82	1.85	
O	Average		1.4	1.53	1.50	1.58	1.40	1.51	**1.49**	1.69	1.74	**1.72**
=O	Carbonyl	O						1.49		1.66	1.70	
-OH	Hydroxyl	OD2,OE2,OG,OH						1.54		1.72	1.77	
-O	Carboxylate	OD*,OE*						1.49		1.66	1.70	
S	Average				1.80	1.81		1.91	**1.84**	1.99	1.99	**1.99**
-S-	Disulfide	SG						1.94		1.99	1.97	
-SH	Sulphydryl	SG,SD						1.88		1.99	2.00	

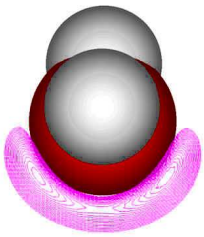
[A] Pauling, L. (1967) The Chemical Bond. Cornell Univ. Press, Ithaca, New York.
[B] Kitaigorodsky, A.I. (1973) Molecular Crystals and Molecules. Physical Chemistry 29,1-.
[C] Bondi, A. (1968) "Molecular Crystals, Liquids and Glasses." Wiley, New York.
[D] Rowland, R.S. and Taylor, R. (1996) Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van Der Waals Radii. J. Phys. Chem. 100, 7384-7391.
[E] Gavezzotti, A. (1983) The Calculation of Molecular Volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity J. Am. Chem. Soc. 105, 5220-5225.
[F] Li, A-J. and Nussinov, R. (1998) A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation , and docking. Proteins 32, 111-127.
[G] Jorgensen, W.L. and Tirado-Rives, J. (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110: 1657-1666.
[H] MacKerell, Jr., A. D. Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. (1998) J. Phys. Chem. B102, 3586-3616.

The above R_{min} are for use in calculating the interactions between two atoms that are NOT connected by bonds. The very notion of using radii to characterize the electron cloud around an atom is a simplification. Below is a picture of the electron density around atoms in a water molecule. In reality, the repulsive part of the Lennard-Jones potential would be governed by the approach of these *electron density clouds*, not necessarily by the approach of *atom centers*. (Actually, it is the Pauli exclusion of electron spins in approaching atomic orbitals that cause the repulsion but the idea is the same).





The attractive part of the L-J potential would be governed by the electrostatic fields of interacting molecules, not directly by the distance between atom centers. Below is a picture depicting the electrostatic potential of water (in pink and contoured at some arbitrary value to look appropriate). Molecular **electric fields** are highly anisotropic and certainly not well described by the radius around a point in space.



e. Electrostatic Interactions

We have talked about the fact that van der Waals attractive forces are due to the *fluctuating* dipoles that are present in all atoms – even neutral atoms. Van der Waals interactions are therefore electrostatic in nature but we quantify them using the L-J potential. We make a somewhat artificial division and consider the electrostatic interaction between *static* charges on atoms as being in a different category called "**electrostatic**" interactions. These interactions are described as being between atoms that have a "permanent" full or partial charge. Full charges are considered to be on groups like the protonated amine of the lysine side chain ($-NH3^{+}$) or the deprotonated carboxylate of the glutamate side chain ($-COO^{-}$). Partial charges are considered to be the permanent partial separation of charge such as found in the water molecule modeled above. Here the oxygen has a partial negative charge of –0.834 e [elementary charge (e) = 1.6×10^{-19} coulombs (C)] and the hydrogens each have a partial positive charge of +0.417 e.

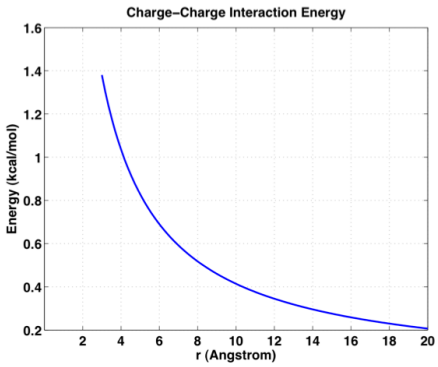
There are two major ways that electrostatic interactions are calculated with empirical force fields: Coulomb's law and Ewald sums. Although more accurate methods exist they are too time consuming to be practical with today's computers.

i. Coulomb's Law

$$\mathbf{E = -q_i q_j / 4\pi D_{ij}}$$

where **q** is the charge on the respective atom, and **D** is the dielectric constant or permittivity of the medium between the charges (relative to a vacuum). The value of **D** ranges from 80 in water to approximately 4 in some regions of a protein core, but force field parameters are usually calibrated to use **D**=1. Notice that the effect of electrostatic

interaction extends beyond the effect of van der Waals interactions, ($1/r$ *vs* $1/r^6$ dependence). Below is a plot of the repulsive Coulombs energy between two same-charge species in water (attractive between opposite charges would be a reciprocal curve). At about 7 Å the energy of interaction becomes less than RT (~0.6 kcal/mol),



ii. Ewald sums

Rather than summing the pairwise electrostatic interactions as in the Coulombic equation above it is possible to simultaneously sum the interactions of a charge with all other charges in a periodic system using Fourier transforms. The summation is done in reciprocal space and when the inverse transform is carried out (with important corrections) the real space electrostatic interaction energy for a charge interacting with *all* of its neighbors is obtained. This works because information on every data point is contained in every term of the FT.

The Ewald sum method is complicated and beyond the scope of this course. If you wish to learn more about it a good description may be found in Allen and Tildesley, page 154 ff (this reference is listed on the course website).

4. Energy scale.

Interaction	Energy (kcal/mol)
van der Waals in water	-0.1
van der Waals in vacuo	-0.3
Hydrogen bond in water	-1
Hydrogen bond in vacuo	-5
Torsion barrier – single bond	+3
Torsion barrier – double bond	+20
Bond stretch by 0.1Å	+2.5
Bond angle change 10°	+2.5
Bond breaking	+100

- van der Waals interactions are weak compared to thermal energy (RT = 0.6 kcal/mol) but can be effective in large numbers such as in the interiors of proteins or membranes where they may act cooperatively.
- Hydrogen bonds are energetically stronger than VDW interactions but are found

only between certain atom types. These bonds are also important both for protein stability and for defining allowed protein conformations.

- The torsion barrier around a single bond is mostly due to the energy to bend bond angles slightly to relieve van der Waals clash and allow the dihedral angle change to occur.
- Bond lengths are allowed to stretch only a small amount but this effect may be cumulative. If all the bonds in a polypeptide backbone stretch by 0.05Å the displacement after 100 residues could be 5Å. The same cumulative effect is present due to backbone angle changes. Some protein flexibility is in bond lengths and angles – not all flexibility is due to bond rotations. Proteins are more like jello than hard solids in this regard.

5. How to calibrate force field parameters?

Where do all the force field parameters ($b, K, \theta, \delta^+, \delta^-, \sigma, \varepsilon$) for each atom or bond type come from? We previously said that these are empirically derived. More specifically, a partial list of sources for the parameter values is:

- **Bond lengths and angles** – small molecule X-ray crystallography
- **Bond spring constants** - IR spectroscopy, quantum mechanical calculations
- **Partial charges** – Assume TIP3 is correct, and then trial and error simulations with water and a small molecule. Iterate parameter values of the small molecule to agree with quantum mechanical calculations and experimental gas-phase dipole moments. (So, one should always use the water model that the force field is calibrated to!).
- **Lennard-Jones parameters** - Trial and error in combination with the partial charge calibration to agree with quantum mechanical calculations and experimental data such as **radial distribution functions**. The following figure is from Jorgensen et al. JACS 106, 6638, 1984 in which the authors calibrate the Lennard-Jones parameters for carbon atoms of the type found in ethane.

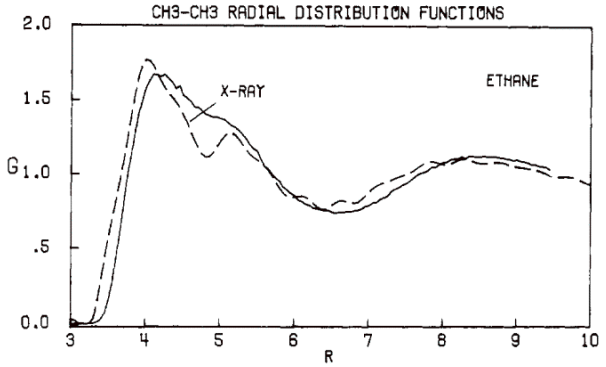


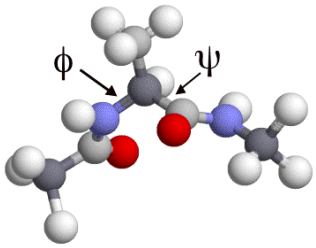
Figure 10. CH_3–CH_3 radial distribution functions for liquid ethane. Computed results (solid curve) are at –89 °C; experimental data (dashed curve) are at –92 °C. Distances are in ångströms thoughout.

6. Comparison of force fields.

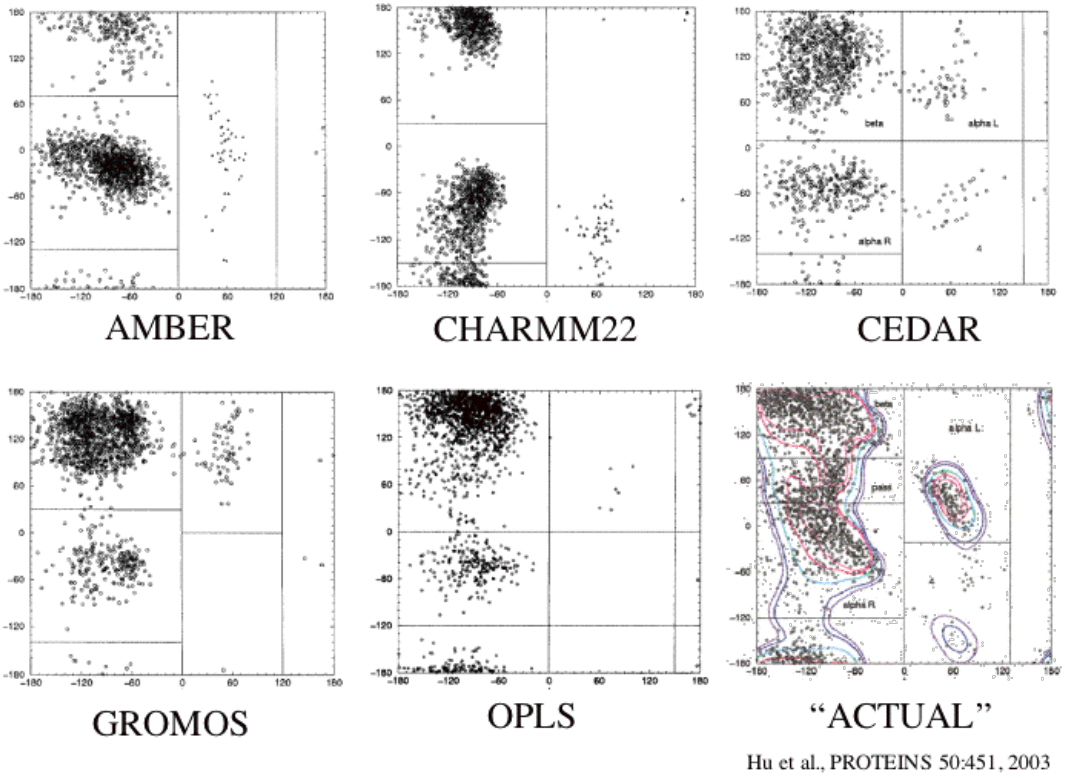
How good are current empirical force fields? It depends on what you want to simulate. To investigate small amplitude dynamics of a *folded* protein (i.e. dynamic flexibility as we will do in the lab) the force fields agree pretty well and are assumed to be adequate. But for simulation of large conformational fluctuations an experiment several years ago

demonstrates that the force fields are not adequate.

Molecular dynamics simulations of a simple model of the polypeptide backbone were done with several different force fields. The distributions of conformations (ϕ, ψ values) were then compared. The experimental molecule was N-acetyl-alanine-N'-methylamide, commonly referred to as alanine dipeptide. The structure is shown in the following figure.



The distributions of ϕ, ψ calculated using the various forcefields were compared to the distributions of ϕ, ψ calculated using quantum mechanical methods (the dots in the plot below labeled "ACTUAL") and also to the distributions of ϕ, ψ from the PDB (the contour lines in the plot labeled "ACTUAL"). Clearly none of the force fields give distributions that match the target distributions. The results in this figure are from 2003 and some improvements have been made, but there are still differences between force fields. There is still work to be done on calibrating and improving force field parameters.



Hu et al., PROTEINS 50:451, 2003

7. Improvement of Force Fields.

Two strategies are being used to improve the accuracy of force fields:

1. Add "knowledge based" information in addition to the "physics based" parameters described above. An example of knowledge based information would be biased sampling (MC) or torsional angle energy restraints (MD) on ϕ, ψ values in polypeptides. The parameters for these biases or restraints come from the distribution of ϕ, ψ values in the PDB. In other words – just add restraints that make the simulated conformational ensemble mimic the observed conformational ensemble.
2. Add polarization to the MM model. Instead of modeling an atom by a point charge, allow the separation of partial charges in the atomic model. This increases the computational time but a recent version of a polarizable force field claims to increase computational time only ~3X.

