Introduction to Force Fields

Functions that associate an energy with a given nuclear configuration (thus they rely on the Born-Oppenheimer approximation).

Key issues we will consider:
• What approximations are made in the functional forms used for biomolecular force fields? Special emphasis: electrostatics and polarizability.
• How are the force fields parameterized?
• How are the force fields different?

A piece of advice:
As we delve into more approximate methods, it becomes important to avoid “black-and-white” thinking, e.g., if the force fields are neglecting some key effect, then their results must be all wrong. All force fields make a great many approximations, and the impact of any given approximation is generally highly context-dependent.
Types of force fields

“Small” molecule
- MM2/MM3/MM4, from Allinger and co-workers, is probably the best known.
- Emphasis on high precision reproduction of geometries and dipole moments.
- Condensed phase not generally considered explicitly; i.e., mostly for “gas phase” results. An exception is early development of OPLS, which was initially for small molecules in condensed phase.
- Extensive parameterization for many different classes of organic molecules; huge number of parameters.

Biomolecule
- The big 3 are CHARMM (Karplus), OPLS (Jorgensen), and Amber (Kollman/Case).
- Emphasis on thermodynamic properties in the condensed phase.
- Condensed phase implicitly taken into account in parameterization.
- Parameterization is tricky and based mostly on small molecule model systems; combination of experimental data and QM.
- Functional forms are simpler than for small molecules; e.g., no 3-body effects.

These two worlds are increasingly merging. Macromolecular force fields increasingly treat arbitrary small molecules, e.g., for protein-ligand simulations. MMFF (Halgren) is more of a mix as well.
History of Biomolecular Force Fields

1970s: First proof of concept papers.
1980s: Foundations for most modern work (big name force fields, MD programs). Scheraga, Karplus, Kollman, etc.
1990s: Lots of commercialization.
Today: Several thousand published papers using molecular dynamics on biomolecules. Variety of relatively easy-to-use programs.

Comments:
• It’s interesting to compare the evolution of MD to the evolution of quantum chemistry. Over roughly the same time frame, QM methods evolved rapidly and in parallel with MD methods, both being driven by increases in computer power. Both fields have, to differing extents, become victims of their own successes. As the methods became accepted, the journals became clogged with papers running QM or MD packages on various systems, and there was some backlash.
• With certain notable exceptions, the field of ligand docking largely abandoned the force-field approach for many years, although there are signs that is changing.
• The field of homology modeling likewise largely abandoned MD/force field-based approaches, after a number of failures.
General Functional Form of All-Atom Macromolecular Force Fields

Bonds \( k_r (r - r_0)^2 \)

Angles \( k_\theta (\theta - \theta_0)^2 \)

Torsions \( \sum_n k_n (\cos n \phi) \)

Nonbonded:

Lennard-Jones \( \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \)

Electrostatic \( \frac{q_i q_j}{r_{ij}} \)

Sources of parameters:

- Gas-phase QM
- Macroscopic properties *via* liquid state simulation, e.g., density, heat capacity, compressibility
- Spectroscopic and crystallographic data (small molecules)
- Guessing!

*Note that there is no special term for hydrogen bonds: treated as simple electrostatics.*

Early force fields did not explicitly treat all hydrogens (especially aliphatic ones): “united atom” force fields. But all modern force fields explicitly treat all atoms.
Covalent Terms

Obviously, each of these terms is a gross approximation
- Bonds: Stretch term of course should be anharmonic to take into account bond dissociation and chemistry. In practice, to get chemistry accurate, generally need to combine QM and MM methods (more later).
- Angles: Anharmonic, but in different ways.
- Torsions: The use of a Fourier expansion (generally 3 terms) is likewise simplistic. [Recent work by Brooks has used a very flexible spline-based functional form for reparameterization of CHARMM torsions.]

Where do the parameters come from?
- Bonds/angles: Gas phase spectroscopy, small molecule crystal structures.
- Torsions: generally gas phase quantum mechanics.

Do these terms even matter?
- Bonds/angles: These are high frequency modes, and generally not too important for thermodynamic quantities we want to calculate (or kinetics on long timescales).
- Torsions: These matter quite a bit, and have been a source of many problems in the past, e.g., tendency of Amber to overstabilize helices. Recent reparameterization has helped for backbones.
Improper Torsions

- These terms have the same functional form as the “proper” torsions, but do not represent a dihedral angle.
- Rather, they represent a deviation from planarity of a set of 4 atoms. An important example is the peptide bond.
- In the absence of these terms, there would be no restoring force for out of plane motions.
- In the original united atom force fields, similar terms were used to maintain the geometry at stereocenters.
A case study from my own work:
Refinement of Asp side chain torsional parameters

Improving fit to high level, gas phase QM improves accuracy of protein side chain prediction
## Non-Covalent Forces

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Force Law</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge-charge</td>
<td>Na$^+$ Cl$^-$</td>
<td>$1/r$</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>Charge-dipole</td>
<td>Na$^+$ OH$_2$</td>
<td>$1/r^2$</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>Dipole-dipole</td>
<td>H-Cl</td>
<td>$1/r^3$</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>Charge-induced dipole</td>
<td>Na$^+$ CH$_4$</td>
<td>$1/r^4$</td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>Dipole-induced dipole</td>
<td>HCl, Cl-H</td>
<td>$1/r^6$</td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Induced dipole-induced dipole</td>
<td>CH$_4$, CH$_4$</td>
<td>$1/r^6$</td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Hydrogen Bond</td>
<td>H$_2$O</td>
<td></td>
<td><img src="image7.png" alt="Image" /></td>
</tr>
</tbody>
</table>

“van der Waals” dispersion forces; deviations from ideal gas behavior
What pairs of atoms count as “nonbonded”? 

That is, for which pairs of atoms should we calculate LJ and Coulomb forces?

• Clearly not bonded pairs, or pairs separated by a bond-angle.
• Pairs of atoms separated by more than 3 bonds should clearly be counted as non-bonded.
• But what about pairs separated by exactly 3 bonds, i.e., those defining the ends of a dihedral angle? This is kind of a gray area.
• Macromolecular force fields generally calculation LJ/Coulomb forces for these “1-4” pairs, but scaled down by some factor. Generally a factor of 2, although AMBER went to a scaling factor of 80%.
• Physically what we care about is reproducing the potential energy as we rotate the dihedral angle. The generic 3-term Fourier expansion, by itself, generally does not do an adequate job, but adding in the nonbonded part with a scaling factor improves agreement.
Van der Waals interactions

Long-Range: Dispersion Forces
- Quantum mechanical effect
- Can be described classically as a spontaneously induced dipole-dipole interaction
- As $r \to \infty$, the interaction scales as $1/r^6$
- Magnitude of force: obviously depends strongly on distance; generally small relative to $kT$. But it adds up ($N^2$ interactions in protein).

Short Range: Strong Repulsion
- Direct consequence of Pauli exclusion principle
- Formally increases exponentially with decreasing internuclear separation
- However, frequently modeled as $1/r^{12}$ (efficiency!)

Put them together: Lennard-Jones formula
- $A/r^{12} - C/r^6$
- Narrow, rather shallow minimum at the sum of the “VDW” radii
Charges from Molecular Electrostatic Potential

Electrostatic potential (ESP) from QM calculation:

$$\phi(\vec{r}) = \phi_N(\vec{r}) + \phi_e(\vec{r}) = \sum_{\alpha} \frac{Z_\alpha}{|\vec{r} - \vec{R}_\alpha|} - \int \frac{d\vec{r}' \rho(\vec{r}^{'})}{|\vec{r}' - \vec{r}|}$$

Basic idea: Fit this quantity with point charges.

This idea has been elaborated by a number of workers:

- Cox and Williams (1981). Introduce least-squares fitting of point charges to ESP, on grid.
- CHELP. ESP on spherical shells surrounding atoms. Also, used Lagrange multiplier approach instead of iterative least-squares (efficiency).
- CHELPG. CHELP, except on a grid instead of spheres. Remove problem with CHELP of not satisfying rotational invariance.
- RESP (Kollman). “Restrained electrostatic potential fit”. Apply restraints to partial charges on some atoms. Removes problem of erratic charges for buried atoms.

There are other, more heuristic methods of assigning partial charges, such as the Gasteiger method, based on electronegativity arguments. Such methods are appropriate primarily when speed is a major concern.
Beyond the Atom-Centered Point Charge Approximation

There is no question that adding off-nucleus point charges can improve fit to QM.
• Various researchers have added, e.g., point charges to represent “lone pairs”.
• Off-atom point charges can also help to reproduce pi-stacking and cation-pi interactions.
• Still, these more complicated charge distributions have not been widely adopted.

Polarizability
• Electron distribution responds to environment!
• Important contribution to, e.g., hydrogen bond strengths.
• Contribution to dielectric response: even purely nonpolar solvent have dielectric ~2.
• Treatment of polarizability – and understanding its significance – is still in its infancy, and much work is being done.
• While we know it is an important physical phenomenon, relevance to biology still incompletely understood.
• We’ll talk more about polarizability in the next lecture.
How to deal with partial charges in condensed phase?

- In the ESP approach, we fit partial charges to *gas phase* QM.
- But in the condensed phase, the effective partial charges generally need to be increased to account for bulk polarizability effects. That is, most polar groups find a way to make favorable hydrogen bonds, which polarize the electron distributions (typically 10-20% increase in dipole moments).
- Example: Water dipole in gas phase is 1.84 D, but the average effective dipole in bulk water is more like 2.5 D.
- The Amber approach has been to use HF 6-31G* to determine the electrostatic potential. The rationale is that this relatively small basis set “fortuitously” predicts dipole moments to be too large relative to gas phase data. But this is what you want for the condensed phase, at least qualitatively.
- There are modern methods for including the effects of bulk water in QM calculations (i.e., combining them with continuum electrostatics, which we will discuss soon), and these increase the bond dipoles.
- The OPLS approach has been to use condensed phase properties like enthalpy of vaporization to help fit the partial charges; in principle, this takes into account bulk polarization.
How to deal with instantaneous/environment specific polarizability?

The extent of polarization of an atom in a macromolecule will depend on its environment. For example, the “hydrophobic core” of a protein is a very different environment that the solvent-exposed exterior. Local details also matter, such as the identity and orientation of hydrogen bonding partners. How do we model this?

For a single atom, the polarizability is represented by a single number, which can be determined experimentally or from QM calculations:

$$\vec{\mu}_{\text{ind}} = \alpha \vec{E}$$

For a molecule, the situation is more complicated; polarization can be anisotropic, and the polarizability is now a tensor:

$$\vec{\mu}_{\text{ind}} = \alpha \vec{E}$$

In practice, we must model this complicated response using relatively simple functional forms. There are two primary methods in use for modeling polarization in a force field:

1. Fluctuating charge (FlucQ).
2. Polarizable point dipole.
• Imagine charge having the ability to move from an atom to its bonded neighbors, such that there is no change in total charge.
• Based on the principle of electronegativity equalization, which in turn is founded on the theory of DFT. The instantaneous electronegativity is the sum of the intrinsic electronegativity and the electrostatic potential, due to surroundings. The energy of transferring an amount of charge between sites \(i\) and \(j\) is, to second order:

\[
E(q_{ij}) = \chi_{ij} q_{ij} + \frac{1}{2} J_{ij} q_{ij}^2
\]

The parameters \(\chi\) and \(J\) can be related to the electronegativity and “hardness” of the individual atoms; in practice, they can be treated as adjustable parameters.

• The correct amount of charge “fluctuation” is found by minimizing the overall energy, which includes both the simple Coulombic interactions and the contribution from the polarization.
• A practical method of solving this is the method of Langrange multipliers. The fluctuating charges are treated as having fictitious (and small) masses. This represents a direct analogy to the Born-Oppenheimer approximation: the electronic degrees of freedom quickly “equilibrate” to new nuclear configuration.
Fluctuating Point Dipoles

Conceptually pretty straightforward: Assign an isotropic polarizability to each atom, and thus allow a point dipole to be created in response to the electric field at that atom. The induced dipoles themselves modulate the electric field, and so you have to solve for the induced dipoles iteratively and self-consistently.

Comparing point dipoles and FlucQ:
• FlucQ can be very efficient, perhaps just 10% greater expense than fixed charge.
• One deficiency of FlucQ is that polarization happens only along bonds, which is not sufficient for some cases. Consider water: FlucQ allows polarization only in the plane of the water, but in fact the polarizability of water is nearly isotropic (i.e., as much polarization out of plane as in it).
• The two methods can be combined, as in the polarizable OPLS force field (Stern, Kaminski, and Friesner).

What is the current state of polarizable force fields?
• They exist, sort of, for OPLS, AMBER, CHARMM. But still under testing and development. Not a whole lot published, really.
• Computational expense is typically 50% (or more) greater than fixed charge.
• Many polarizable water models published, but still not widely used.
Validating Force Fields

• The key issue is transferability. Parameterization must be performed using relatively small model compounds. Gas phase data used for some parameters. How do the parameters hold up on a solution phase macromolecular simulation?
• Historically, the primary method of validating the parameters would be to either minimize or run molecular dynamics on a crystal structure; the deviation from the experimental structure would be taken as a measure of force field adequacy.
• Decoy studies are a popular (but flawed) modern method; is it possible to distinguish native conformations from non-native?
• These methods are a fine first test, but obviously have a number of limitations. Among other things, fine details of crystal structures can be related to crystal packing, which is generally not considered in simulations. NMR structures are rarely of high enough resolution to be useful for these purposes. Also, it is difficult to determine exactly what is wrong with the force fields.
• In my own work, I have championed the use of side chain and loop prediction as a means of validating, comparing, and improving force fields, with crystal packing explicitly included for a fair comparison with the experimental data.
• One serious problem with validating force fields, which is perhaps not as widely appreciated as it should be, is incorrect protonation states. We generally want partial charges to be accurate to about 0.1 units; if a protonation state is wrong, we’re off by an entire unit charge, which can have dramatic and long-range effects.
Process of Parameterization

• ... is extremely labor intensive. For the most part, it continues to be an “artisanal” process; there are many judgment calls, in practice.
• A key decision is how many “atom types” to introduce. In other words, are all carbon atoms treated the same? All carbonyl carbon atoms treated the same? Or are there different parameters for ketones, aldehydes, carboxylates, etc.?
• The AMBER strategy has historically been to use a relatively small number of atom types; for AMBER and other force fields, however, the number of atom types has grown significantly over the years, and continues to grow. Most modern force fields want to treat not only amino acids and nucleic acids but also arbitrary ligands; this is generally incompatible with using a small number of atom types.
• For standard amino acid residues and nucleic acids, the atom types and partial charges can of course be defined in advance, and the various parameters extensively checked. But for arbitrary ligands, need reasonably sophisticated technology to identify atom types and assign the “best” parameters.
• Co-variance of parameters is a serious problem! [i.e., many solutions exist with similar fit to the data used for parameterization]
So what’s the best force field?

- That’s largely an ill-posed question, because the answer is so context-dependent.
- There has been a cottage industry of comparing different force fields on different types of problems (peptide folding, decoys, RNA/DNA, etc.). The answers are generally mixed, although it is clear that AMBER has historically had some problems with backbone torsional parameters, possibly fixed in recent versions.
- Even within a single force field, there are so many revisions, and performance is different for each one.
- The answer is also complicated by the fact that you have to chose a model for the solvent, and this can affect the results as much as the model for the solute. Keep in mind that force fields have generally been parameterized with a particular solvent model in mind.
- Finally, it is important to point out that there has been a fair amount of cross-fertilization among the major force fields. OPLS originally took covalent parameters from AMBER, AMBER later borrowed some parameters from OPLS, etc. In general, there appears to be some level of convergence among the force fields, although they continue to be developed and used by largely distinct groups of researchers.
Hierarchy of Methods So Far

HF + CI

HF      DFT

Semi-empirical

Force fields

single-point calculations on small model systems

geometry optimization on small/intermediate model systems

extensive sampling of macromolecular systems

Comments:
• Force fields (of the usual type) cannot be used to study chemistry (bond-breaking reactions, e.g., in enzymes). Must use QM, or combine QM/MM.
• Force fields work well (or at least as well as they can) at ambient temperatures and equilibrium conditions.
• We have not talked about sampling yet, i.e., exploring the energy surface in some defined way. There is a trade-off between accuracy of the energy function (and thus its expense) and the thoroughness of sampling.