Subject of the Lecture:

- Conceptual basis for the development of force fields.
- Implementation/validation
- Water - a worked example
- Extensions - combining molecular mechanics and quantum mechanics (QM/MM)
Currently, complete quantum mechanical solution to most problems of chemical or biological interest is not possible.
Difficulties are compounded by the fact that statistical and dynamical aspects of the problem must be included. In fact, they are the essence of biology.

$$\Omega_{\text{tot}}(N, V, E) = \frac{1}{h^{3N} N!} \int \int_{V_N} \delta[H(q^N, p^N) - E] dq^N dp^N$$
Schrödinger’s equation:

\[
\frac{\hbar}{i} \frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t)
\]

Newton’s equations of motion

\[
F(x,t) = m \frac{d^2x}{dt^2} ; \quad F(x,t) = -\frac{\partial V(x,t)}{\partial x}
\]

Can we consider molecular systems as classical?
Conceptual Basis

Quantum mechanics provides clues to the form of the potential function.

We will analyze the problem stepwise.

Consider two interacting molecules A and B:
$$H = H_A + H_B + H_{int}$$
Conceptual Basis - Electrostatics
(Rayleigh-Schrodinger perturbation theory)

\[ \varphi(r) = \frac{q}{4\pi\varepsilon_r} \sum \left( \frac{r}{R} \right)^n P_n(\cos\theta) \]

Multipole expansion - charge density is represented as a sum of point multipoles (charge, dipole, quadrupole)
Conceptual Basis (2) - Electrostatics

- $\sim R^{-1}$ – charge-charge
- $\sim R^{-2}$ – charge-dipole
- $\sim R^{-3}$ – dipole-dipole,
- - charge-quadrupole
- ...

$q_A = q_B = 0$
In the absence of charge transfer the first non-vanishing term is dipole
Conceptual Basis - dispersion energy (electron correlation)

Dispersion energy is always negative

$$E_{\text{disp}} \sim R^{-6} (+ R^{-8} + ...)$$
Conceptual Basis - short-ranged potential

- Beyond the R-S perturbation formalism ("symmetry-adapted" theory)
- Arises from Pauli’s exclusion principle
- In the “first-order” approximation proportional to the electron overlap
decays exponentially with the distance
Always positive
van der Waals potential
(electron correlation)

Buckingham potential:
\[ E_{vdw} = -Ar^{-6} + B\exp(-\alpha r) \]

Lennard-Jones potential:
\[ E_{vdw} = -Ar^{-6} + Br^{-12} \]

\[ E_{vdw} = 4\varepsilon[-\frac{\sigma^6}{r^6} + \frac{\sigma^{12}}{r^{12}}] \]

\[ r_{\text{min}} = 2^{1/6}\sigma; \quad E_{vdw}(\text{min}) = \varepsilon \]
Van der Waals potential (cont.)

\[- \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}\]

Repulsion regime

van der Waals attraction regime

Optimum energy
Van der Waals potential (cont.)

\[-\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}\]

A = 0.5  B = 1.0

A = 1.0  B = 5.0
Angle-bending
Stretching or bending

\[ k_\theta (\theta - \theta_0)^2 \quad \text{or} \quad k_b (r - r_0)^2 \]
Torsion

\[ A [1 + \cos(n\tau - \phi)] \]
Torsions (cont.)

\[ A [1 + \cos(n\tau - \phi)] \]

- \( A = 2.0, n = 2.0, \phi = 0.0^\circ \)
- \( A = 1.0, n = 1.0, \phi = 90.0^\circ \)
- \( A = 1.0, n = 2.0, \phi = 0.0^\circ \)
SUMMARY

• $E = (E_{el} + E_{vdw}) + (E_{stretch} + E_{bend} + E_{tors} + E_{it})$

• Electrostatic energy is represented using multipole expansion.

• In general, polarization energy is markedly smaller than unperturbed energy.

• van der Waals energy represents electron correlation. The dispersion term is always negative whereas short-range energy is always repulsive. Taken together, $E_{vdw}$ has a minimum.

• The remaining terms describe covalent bonding.
Implementation - a “Lego” force field

Different atom types have different parameters. Same atom types have the same parameters.
• 'Ca': carbon in O-C=O acid
• 'C k': carbon in C=O not bonded to N, and not an acid group
• 'C n': carbon in N-C=O
• 'C*': aromatic C in 5-membered ring next to two carbons
• 'CAxna': special parameters for nucleotide bases. aromatic C in DAP C2,C3,C4 pyridine; Cytosine C4,C5,C6
• 'CBxna': special parameters for nucleotide bases. Adenine C4,C5,C6; Guanine C2,C4,C5
• 'CB': aromatic C at juncton of 5-membered and 6-membered rings
• 'CKxna': special parameters for nucleotide bases. Adenine C8 Guanine
• 'CM': alkene sp2 carbon (non-aromatic)
• 'CMxna': special parameters for nucleotide bases. Carbon in Uracil C5,C6
• 'CO': sp3 anomereric carbon (bonded to ether O and alcohol O); this type shows up in carbohydrates
• 'CQ': sp2C in 6-membered ring between deprotonated N's
• 'CR': aromatic C in 5-membered ring next to two nitrogens
• 'CT': aliphatic sp3 hybrid carbon
• 'CV': aromatic C in 5-membered ring next to C and deprotonated N
• 'CW': sp2 aromatic C in 5-membered ring next to C and NH
* 'O' : oxygen in C=O, not an acidic site
* 'O2' : oxygen double bonded to carbon in COO- or COOH
* 'OHa' : oxygen bonded to H in RCOOH
* 'OHm' : oxygen bonded to H in a mono-alcohol
* 'OHp' : oxygen bonded to H in polyols or phenol
* 'OS' : sp3 O in an ether or acetal
We need to assign and validate parameters

- Charges on atoms (for each atom type)
- Other electrostatic parameters (if needed)
- Van der Waals parameters (for each pair of atom types)
- Force constants and equilibrium distances (angles) for stretching (bending)
- Parameters for torsions
We need to assign and validate parameters

- Theory or experiment?
- Isolated molecules or dense phases?
- Large or small molecules?
- Structure? Spectroscopy? Thermodynamics?
- Pure substances or mixtures?

Every force field should be well balanced
Typical sources/validation of other potentials

- Charges - electrostatic potential
- Van der Waals parameters - crystallography
- Stretching & bending parameters - spectroscopy
- Torsional parameters - quantum mechanics, NMR spectroscopy
Different Force Fields:

- **AMBER** (Assisted Model Building with Energy Refinement).
- **CHARMM** (Chemistry at HArvard using Molecular Mechanics).
- **GROMOS** (GROenigen MOlecular Simulation)
- **OPLS** (Optimized Parameters for Large-scale Simulations)
- **MMFF** (the Merck Molecular Force Field)
- Other (MM3, Tripos,…)
Who is the fairest of them all?
Force Fields - a worked example: water models

- Water is **the** biological solvent
- Water has many unusual properties
- Bulk, bound, crystal and droplet water is important
Different representations of water
## Models of water:

<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th>(\sigma) Å</th>
<th>(\varepsilon) kJ mol(^{-1})</th>
<th>(l_1) Å</th>
<th>(l_2) Å</th>
<th>(q_1)</th>
<th>(q_2)</th>
<th>(\theta) °</th>
<th>(\phi) °</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD [511]</td>
<td>-</td>
<td>3.016</td>
<td>15.319</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>109.47</td>
<td>109.47</td>
</tr>
<tr>
<td>SPC [94]</td>
<td>a</td>
<td>3.166</td>
<td>0.650</td>
<td>1.000C</td>
<td>-</td>
<td>+0.410</td>
<td>-0.8200</td>
<td>109.47</td>
<td>-</td>
</tr>
<tr>
<td>SPC/E [31]</td>
<td>a</td>
<td>3.166</td>
<td>0.650</td>
<td>1.000C</td>
<td>-</td>
<td>+0.4238</td>
<td>-0.8476</td>
<td>109.47</td>
<td>-</td>
</tr>
<tr>
<td>SPC/HW (D(_2)O) [220]</td>
<td>a</td>
<td>3.166</td>
<td>0.650</td>
<td>1.000C</td>
<td>-</td>
<td>+0.4350</td>
<td>-0.8700</td>
<td>109.47</td>
<td>-</td>
</tr>
<tr>
<td>TIP3P [180]</td>
<td>a</td>
<td>3.15061</td>
<td>0.6364</td>
<td>0.9572</td>
<td>-</td>
<td>+0.4170</td>
<td>-0.8340</td>
<td>104.52</td>
<td>-</td>
</tr>
<tr>
<td>PPC 1, 2 [3]</td>
<td>b</td>
<td>3.23400</td>
<td>0.6000</td>
<td>0.943C</td>
<td>0.06</td>
<td>+0.5170</td>
<td>-1.0340</td>
<td>106.00</td>
<td>127.00</td>
</tr>
<tr>
<td>TIP4P [180] 10</td>
<td>c</td>
<td>3.15365</td>
<td>0.6480</td>
<td>0.9572</td>
<td>0.15</td>
<td>+0.5200</td>
<td>-1.0400</td>
<td>104.52</td>
<td>52.26</td>
</tr>
<tr>
<td>TIP4P-FQ [197]</td>
<td>c</td>
<td>3.15365</td>
<td>0.6480</td>
<td>0.9572</td>
<td>0.15</td>
<td>+0.63(^1)</td>
<td>-1.26(^1)</td>
<td>104.52</td>
<td>52.26</td>
</tr>
<tr>
<td>SWFLEX-AI 2 [201]</td>
<td>c</td>
<td>four terms used</td>
<td>0.968(^1)</td>
<td>0.14(^1,3)</td>
<td>+0.6213</td>
<td>-1.2459</td>
<td>102.7(^1)</td>
<td>51.35(^1)</td>
<td></td>
</tr>
<tr>
<td>COS/G3 [704] 11</td>
<td>c</td>
<td>3.17459</td>
<td>0.9445</td>
<td>1.000C</td>
<td>0.15</td>
<td>+0.450672</td>
<td>-0.901344</td>
<td>109.47</td>
<td>-</td>
</tr>
<tr>
<td>TIP5P [180]</td>
<td>d</td>
<td>3.12000(^6)</td>
<td>0.6694(^9)</td>
<td>0.9572</td>
<td>0.70</td>
<td>+0.2410</td>
<td>-0.2410</td>
<td>104.52</td>
<td>109.47</td>
</tr>
<tr>
<td>POL5/TZ 2 [256]</td>
<td>d</td>
<td>2.98374(^4)</td>
<td>0.9572</td>
<td>0.5</td>
<td>varies(^5)</td>
<td>-0.42188</td>
<td>104.52</td>
<td>109.47</td>
<td></td>
</tr>
<tr>
<td>Six-site [491]</td>
<td>c/d</td>
<td>3.115(<em>{\text{OO}}) 0.673(</em>{\text{HH}})</td>
<td>0.715(<em>{\text{OO}}) 0.115(</em>{\text{HH}})</td>
<td>0.980</td>
<td>0.8892(<em>{\text{L}}) 0.230(</em>{\text{M}})</td>
<td>+0.477</td>
<td>-0.044(<em>{\text{L}}) -0.866(</em>{\text{M}})</td>
<td>108.00</td>
<td>111.00</td>
</tr>
</tbody>
</table>
Parameterizing a water model

- Simple averages:
  - density ($\rho$)
  - enthalpy of vaporization ($\Delta H_{vap}$)
  - structure (O-O RDF)
  - dipole moment

- Higher-order averages:
  - heat capacity ($C_p$)
  - dielectric constant ($\varepsilon$)
  - isothermal compressibility ($\kappa$)
  - coefficient of thermal expansion ($\alpha$)

- Kinetic observables:
  - diffusion constant (D)
  - rotational autocorrelation ($\tau_{rot}$)
  - hydrogen bond lifetimes ($\tau_{HB}$)

A particular challenge is getting the temperature dependence of these right, e.g., the density maximum of water around 4 C.
<table>
<thead>
<tr>
<th>Model</th>
<th>Dipole moment</th>
<th>Dielectric constant</th>
<th>Self diffusion, $10^{-5}$ cm$^2$/s</th>
<th>Average configurational energy, kJ mol$^{-1}$</th>
<th>Density maximum, °C</th>
<th>Expansion coefficient, $10^{-4}$ °C$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP4P-FQ</td>
<td>2.64 [197]</td>
<td>79 [197]</td>
<td>1.93 [197]</td>
<td>-41.4 [201]</td>
<td>+7 [197]</td>
<td>-</td>
</tr>
<tr>
<td>SWFLEX-AI</td>
<td>2.69 [201]</td>
<td>116 [201]</td>
<td>3.66 [201]</td>
<td>-41.7 [201]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Six-site*</td>
<td>1.89 [491]</td>
<td>33 [491]</td>
<td>-</td>
<td>-</td>
<td>+14 [491]</td>
<td>2.4 [491]</td>
</tr>
<tr>
<td>Expt.</td>
<td>2.65, 3.0</td>
<td>78.4</td>
<td>2.30</td>
<td>-41.5 [180]</td>
<td>+3.984</td>
<td>2.53</td>
</tr>
</tbody>
</table>
The radial distribution function represents the average density as a function of distance from a particular particle. It can be measured experimentally by a variety of means. Many thermodynamic properties can be derived from the RDF.
What’s missing from these water models?

- Polarizability
- Bond flexibility
- Dissociation
- Multibody effects
- Purely quantum effects

<table>
<thead>
<tr>
<th>Dipole Moments</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas phase</td>
<td>1.85 D</td>
</tr>
<tr>
<td>TIP4P</td>
<td>2.18 D</td>
</tr>
<tr>
<td>SPC</td>
<td>2.27 D</td>
</tr>
<tr>
<td>Liquid water</td>
<td>~2.5 D</td>
</tr>
</tbody>
</table>

An ongoing challenge: Water in first solvation shell of protein (or “trapped” in an active site or the interior) is rather different than bulk water.
Hybrid QM/MM

- Combines quantum mechanical and molecular mechanical methods
- Treats just the reacting part of the system quantum mechanically, and uses MM for the surroundings
- Uses a combined Hamiltonian for the system:

\[
\hat{H}_{total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}
\]
An Example - a Diels-Alder Reaction

\[ \text{Cl}_2\text{C} = \text{S} = \text{O} + \text{C} = \text{O} - R \xrightarrow{1E9} \text{Cl}_2\text{C} = \text{S} = \text{O} \]

\[ -R = -(\text{CH}_2)_5\text{CO}_2^- \]
The Diels-Alder cyclo-addition reaction catalyzed by 1E9, a catalytic antibody that was raised against a transition-state analogue compound.
\[ \hat{H}_{\text{eff}} = \hat{H}_{QM/MM} + \hat{H}_{MM/MM} + \hat{H}_{QM/MM} \]

\[ E = \langle \Phi | \hat{H}_{\text{eff}} | \Phi \rangle = \langle \Phi | \hat{H}_{QM/MM} | \Phi \rangle + \langle \Phi | \hat{H}_{QM/MM} | \Phi \rangle + E_{MM/MM} \]

QM/MM Hamiltonian

\[ \hat{H}_{QM/MM} = - \sum_{i,M} \frac{e q_M}{r_{iM}} + \sum_{\alpha,M} \frac{Z_{\alpha} q_M}{R_{\alpha M}} + \sum_{\alpha,M} 4 \varepsilon_{\alpha M} \left\{ \frac{R_{\alpha M}^{12}}{R_{\alpha M}^{12}} - 2 \frac{R_{\alpha M}^{6}}{R_{\alpha M}^{6}} \right\} + \hat{H}_{\text{int-coor}}^{QM/MM} \]

Counters

- \( i \): electrons
- \( \alpha \): nuclei
- \( M \): MM atoms

\( e \): electron charge
\( q_M \): atomic charge on MM atoms
\( Z \): nuclear charge
\( R/r \): distances between particles
\( \varepsilon \) and \( R_{\text{min}} \): well-depth and radius
Application of QM/MM

In solution

Reactants being treated by QM are surrounded by MM waters: no covalent bonds between QM atoms and MM atoms

Enzyme active site

Necessity of including amino acid or coenzyme functional groups in QM region requires treatment of covalent link between QM and MM regions (frontier region). (Reuter et al., 2000)
Treatment of QM/MM boundaries

How do we deal with bonds between the QM and MM regions?
- The valence of the QM region must be satisfied
- MM bond, angle, dihedral terms need a partner atom to act on, in order to maintain the geometry of the system
- QM/MM is often used to simulate a solute quantum mechanically, with explicit solvent treated with MM — in this instance, the problem of QM-MM bonds is avoided
Link atoms

Conventional solution: ‘link atoms’ (usually hydrogen atoms, but sometimes halogens or even methyl groups) are added along the bond $^a$

- The link atom satisfies the valence of the QM region
- The QM atom is used for calculation of all MM bond terms
- For nonbond (electrostatic terms), originally the link atom did not interact with any MM atom (termed a ‘QQ’ link in CHARMM parlance)
- Better properties are usually obtained if the link atom interacts with the entire MM region (‘HQ’ link)
- Poor handling of electron density
Improved bond treatments

- Local Self-Consistent Field (LSCF) \(^a\) uses a parameterized frozen orbital along the QM-MM bond, which is not optimized in the SCF.

- Generalized Hybrid Orbital (GHO) \(^b\) includes the QM-MM orbitals in the SCF.
A simplification

ONIOM \(^a\) divides the system into the ‘real’ (full) system and the ‘model’ (subset) and treats the model at high level, and the real at low level, giving the total energy as

\[ E(\text{high, real}) \approx E(\text{low, real}) + E(\text{high, model}) - E(\text{low, model}) \]

which relies on the approximation

\[ E(\text{high, model}) - E(\text{low, model}) \approx E(\text{high, real}) - E(\text{low, real}) \]

- The ‘model’ system still has to be properly terminated
- Extension to three level systems is relatively straightforward (e.g. \textit{ab initio} core, semi-empirical boundary, MM surroundings)

Other approaches:

- General Valence Bonds
- Rappé-Goddard
- Car-Parrinello
- Path integral methods


For more info e-mail Andrew Pohorille at pohorill@max.arc.nasa.gov