How did it work out?

\[ \frac{\delta I}{\delta c_A} = 0 \ ; \ \frac{\delta I}{\delta c_B} = 0 \implies \text{secular equation} \]
Where were we?

The electronic Schrödinger equation:

$$H\Psi(X,x) = E_{el}\Psi(X,x)$$

$H$ is the electronic Hamiltonian: $T_e + V_{Ne} + V_{ee} + V_{NN}$

$T_e = -1/2 \sum \{d/dr_i^2\}$ kinetic energy of electrons

$V_{Ne} = \sum \sum \{-Z_iA/r_{iA}\}$ electron-nuclear attraction

$V_{ee} = 1/2 \sum \sum 1/r_{ij}$ electron-electron repulsion

$V_{NN} = 1/2 \sum \sum \{Z_iZ_j/R_{ij}\}$ nuclear-nuclear repulsion
Molecular mechanics (MM) or Force field (FF) methods

From the Born-Oppenheimer approximation:
Schrödinger equation for nuclear motion: \((T_N + V_N(X)) \chi = \varepsilon \chi\)

\(V_N(X) = \text{nuclear repulsion} + E_{\text{el}}, \ E_{\text{el}} \text{ is obtained from solving the electronic Schrödinger equation for different nuclear positions}\)

In this potential \((V_N(X))\) the energy \(\varepsilon\) depends only on the nuclear coordinates \((X)\)

**Approximations:**
1. Classical nuclear motion: \(F_i = ma_i \quad F_i = -\partial V_N({x_i})/\partial \ x_i\).
2. \(V_N(X)\) is approximated by a parametrized function: \(E(X)\)

What is the functional form of \(E\)?

Should describe the changes in potential energy when the nuclear coordinates change

- **Notice! Only description of accessible region required**
The force field energy:

**Intra-molecular degrees of freedom**

- **bond stretching**
  - ![Bond Stretching](image)
- **angle bending**
  - ![Angle Bending](image)
- **torsions**
  - ![Torsions](image)

**Inter-molecular degrees of freedom**

- **relative positions and orientations**
  - ![Relative Positions](image)
The force field energy: \[ E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \]

- \( E_{str} \): bond stretching
- \( E_{bend} \): angle bending
- \( E_{tors} \): torsion (bond rotation)
- \( E_{vdw} \): van der Waals interaction
- \( E_{el} \): electrostatic interaction

Non bonded interactions: \( E_{vdw} + E_{el} \)

\( E_{cross} \): couplings between \( E_{str} \), \( E_{bend} \), \( E_{tors} \)

Functional form of each term? (parametrisation: empirical / ab initio)
General purpose force fields:
Molecules are composed of structurally similar units.

Note: Many different atom types, e.g.
C $sp^3$-carbon
C $sp^2$-carbon, alkene
C $sp^2$-carbon, carbonyl
C $sp^2$-carbon, aromatic
C $sp$-carbon
C cyclopropane-carbon
C radical-carbon
etc, etc

Assumption: For a particular pair of atom types
i) the optimal bond distance is the same, and
ii) the energy change upon distortion is the same (force constant)
Similar for bend, torsion etc.
What can force field methods be used for?
1. Determine **structures** (minimize $E_{FF}$ with respect to the nuclear coordinates)
2. Compare **energies** of different structures (more difficult)
3. Simulate systems at finite temperature, free energy
4. Establish rigorous link to statistical mechanics
5. Dynamical information ...................diffusion/reaction rates

Why use force field methods?
1. Relatively **fast** (many medium sized molecules)
2. **Large systems** (e.g. Proteins)
3. Combine with quantum mechanics: **QM-MM**

Limitations of force field methods?
• Accuracy  Predictability  Transferability
• Reactions very difficult
  *(Garbage in – garbage out)*
Many different force field methods exist. They differ e.g. in:

1. Types of interactions included
2. Functional form of each term
3. Inclusion of cross terms
4. Type of information used for fitting the parameters:
   i) empirical  different types of experimental data
   ii) ab initio  data from electronic structure calculations

Different purposes of different methods:

Class I methods: Proteins, DNA etc
Class II methods: Smaller molecules, “high” accuracy
Typical system to have in mind:

http://www.theochem.kth.se/~hakan/
The bond stretch energy

Taylor expansion for the energy:

\[ E_{\text{str}}(R_{AB} - R_{0\ AB}) = E(0) + \frac{dE}{dR}(R_{AB} - R_{0\ AB}) + \frac{1}{2}\frac{d^2E}{dR^2}(R_{AB} - R_{0\ AB})^2 + \ldots. \]

\[ E(0) \equiv 0, \quad \frac{dE}{dR} = 0 \quad \text{(equilibrium)} \]

Second order = harmonic approximation:

\[ E_{\text{str}}(R_{AB} - R_{0\ AB}) = k_{AB} (R_{AB} - R_{0\ AB})^2 \]

Parameters: \( k_{AB} \) : force constant, \( R_{0\ AB} \) : “natural” bond length

Improvements: Third and fourth order terms (more parameters)

Alternative: Morse potential: \( E_{\text{Morse}}(\Delta R) = D_e (1 - e^{-\alpha \Delta R})^2 \) \( (\Delta R = R - R_0) \)

Goes to dissociation energy at large \( R \), parameters: \( D_e, \alpha, R_0 \).
The bending energy

Taylor expansion and harmonic approximation:

\[ E_{\text{bend}}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC} (\theta^{ABC} - \theta_0^{ABC})^2 \]

Parameters: \( k^{ABC} \) : force constant
\( \theta_0^{ABC} \) : “natural” bond angle

Higher terms: mainly cubic

Special problems:
Di-or trivalent central atom: barrier for 180°
Special atom types for small rings
The torsion energy

1. The energy is periodic in $\omega$
2. Low cost for distortion $\Rightarrow$ large distortions

Taylor expansion not useful, use Fourier series (periodic):

$$E_{\text{tors}}(\omega) = \sum_n V_n \cos(n\omega)$$

$n=1$ periodic by $360^\circ$, $n=2$ periodic by $180^\circ$, $n=3$ periodic by $120^\circ$

$V_n$ determines the barrier height

The general expression used (several minima common).

$$E_{\text{tors}}(\omega^{ABCD}) = \frac{1}{2} V_1^{ABCD}(1+\cos(\omega^{ABCD})) + \frac{1}{2} V_2^{ABCD}(1-\cos(2\omega^{ABCD})) + \frac{1}{2} V_3^{ABCD}(1+\cos(3\omega^{ABCD}))$$

For each angle $\omega^{ABCD}$: three parameters $V_1^{ABCD}, V_2^{ABCD}, V_3^{ABCD}$
Electrostatic energy (non-bonded atoms)

The distribution of electrons gives positive and negative parts of the molecules, which interact with each other

1. Assign partial atomic charges, $Q^x$

$$E_{el}(R^{AB}) = \frac{Q^AQ^B}{\varepsilon R^{AB}}$$

$\varepsilon$ can be used to describe the screening by solvent molecules (atoms)

2. Alternative: Assign dipoles ($\mu^x$) to the bonds in the molecule

$$E_{el}(R^{AB}) = \frac{\mu^A\mu^B}{\varepsilon (R^{AB})^3}(\cos \chi - 3\cos \alpha^A \cos \alpha^B)$$

$R^{AB}$ distance, $\chi$, $\alpha^A$, $\alpha^B$ angles between the dipoles.

Charge method most often used (easier to parametrize).

Anisotropy and polarization effects: multipole distributions and induced dipoles
Repulsive + van der Waals energy
non-bonded atoms

1. Repulsive part: overlap of electron clouds (short distance), exchange interaction (Pauli principle, QM)

2. Attractive part: dispersion forces, induced dipoles, quadrupoles etc.

Dipole-dipole: \((R_{AB})^{-6}\)
Dipole-quadrupole: \((R_{AB})^{-8}\) etc

\[ E_{vdw}(R^{AB}) = E_{rep}(R^{AB}) - \frac{C^{AB}}{(R^{AB})^6} \]

\[ E_{rep}(R^{AB})? \]
1. **Lennard-Jones potential**
   
   \[ E_{\text{LJ}}(R) = C_1/R^{12} - C_2/R^6 \]  
   
   \[ E_{\text{LJ}}(R) = \varepsilon [(R_0/R)^{12} - 2 (R_0/R)^6 ] \]  
   
   \( C_1 \) and \( C_2 \) are parameters  
   
   \( \varepsilon \) and \( R_0 \) are parameters

2. **Buckingham (exp-\( R^{-6} \)) potential**
   
   \[ E_{\text{vdw}}(R_a) = Ae^{-BR} - C/R^6 \]  
   
   \( A, B \) and \( C \) parameters

3. **Morse potential**
   
   \[ E_{\text{Morse}}(R) = D(1-e^{-\alpha R})^2 \]  
   
   \( D \) and \( \alpha \) parameters

Short distance: Morse better than exp-\( R^{-6} \) better than Lennard-Jones  
For practical purposes: no big difference in accuracy

**Lennard-Jones potential** computationally most efficient

Some force fields have a special treatment of **hydrogen bonds**
Cross-terms

The cross-terms describe the coupling between e.g. bond-lengths and angles. They can be represented by products of Taylor-like expansions. The most important is:

\[ E_{\text{str/bend}} = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})[(R^{AB} - R_0^{AB}) + (R^{BC} - R_0^{BC})] \]

parameter: \( k^{ABC} \), \( k^B \) or \( k \)

Example:

- \( \theta^{ABC} = 104.5^\circ \) for optimal conformation (opt)
- \( \theta^{ABC} = 90^\circ \) frozen
Polarizable force fields

Water dipole in gas phase 1.89 Debye
In liquid water ~3.0 Debye

Permanent charge distribution + polarizability
Induced dipoles
Interact and needs to be determined self-consistently

Easily divergent
Iterative process

Evaluated against QC calc.
Localized (Wannier) orbitals

Implicit solvent models
Electronic, orientational diffusional polarization
Comparing energies

\[ E_{\text{FF}} = \text{“steric energy + electronic energy”}, \quad \text{arbitrary zero point} \]

Pauli repulsion electrostatic polarization...

Relative energies: \( \Delta E = \Delta E_{\text{FF}} \) if:

i. Different conformations of the same molecule
ii. The same number of structural units

For chemically different molecules: convert to heats of formation:

\[ \Delta H_f = E_{\text{FF}} + \sum_{\text{bonds}} \Delta H_{AB} + \sum_{\text{groups}} \Delta H_G \]

\( \Delta H_{AB} \) and \( \Delta H_G \) are heats of formation parameters fitted to exp. data

Only a few force fields have heat of formation parametrized
## Force field parameters

The number of parameters needed is very large

Example: 71 atom types (MM2 1991)

<table>
<thead>
<tr>
<th></th>
<th>Total number</th>
<th>Actual number</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{vdw}$</td>
<td>142</td>
<td>142</td>
</tr>
<tr>
<td>$E_{str}$</td>
<td>900</td>
<td>290</td>
</tr>
<tr>
<td>$E_{bend}$</td>
<td>27 000</td>
<td>824</td>
</tr>
<tr>
<td>$E_{tors}$</td>
<td>1 215 000</td>
<td>2 466</td>
</tr>
</tbody>
</table>

For each parameter: 3-4 independent exp. data needed

Alternative: QM calculations

**Note:**

i. Parameters only partially transferable between programs

ii. Additional parameters must be determined in the same way as the old ones
Intramolecular - non-bonded energy terms

Include $E_{vdw}/E_{el}$ for atom pairs separated by three or more bonds

$E_{vdw}/E_{el}$ contributions grow as the square of the system size

Compare: $E_{str}/E_{bend}/E_{tors}$ contributions grow linearly with the system size

Use cutoff parameter for non-bonded terms, longer cutoff distance for $E_{el}$ than for $E_{vdw}$

Or in periodic systems: $E_{el}$ : Ewald sum method ($N^{3/2}$)
Fast multipole method (linear)
Force fields for metal complexes

Varying number of ligands

Different geometric arrangements
  e.g. two types of CO ligands: CO$_{ax}$, CO$_{eq}$

Trans effects etc.

Lack of well defined bonds

Lack of experimental data:
  QM calculations used

Force fields for metal complexes defined for special classes of molecules
Some practical aspects

Artificial minima can be obtained

Force fields improve with time

Atom types and bonding usually specified by the user (sometimes a graphical interface exist)

Errors can not be estimated within the model
General purpose potential energy models are used as force fields developed for particular areas of science:

Proteins, Sugars, DNA, metal oxides, solvents, (reactions) etc

functional forms of $E(X)$ differs significantly

(We will look at the Amber force force)

The same force fields can be employed in

Molecular mechanics: static calc., optimisations

Monte Carlo simulations: finite temperature, disorder

Molecular Dynamics simulations: dynamical properties

Potential energy Nuclear forces Hessian
Large models: Hybrid QM-MM methods

Reaction site
Crystal structure of the heme region in Mb compound II at pH 5.2.

Theoretical chemistry can help resolve the link between structure and function.

Hersleth H et al. J. Biol. Chem. 2007;282:23372-23386

©2007 by American Society for Biochemistry and Molecular Biology
Prize in Chemistry for 2013 to Martin Karplus, Michael Levitt and Arieh Warshel

"for the development of multiscale models for complex chemical systems"
**Static QM/MM**
- Reaction site QM + Remaining protein MM + no explicit solvent
- No sampling

**Dynamic QM/MM**
- Reaction site QM + Remaining protein MM + solvent MM
- Molecular dynamics Monte Carlo

**Full ab initio**
- Reaction site + Remaining protein + solvent
- Preferably crystalline
- Molecular dynamics Monte Carlo
Large models are difficult

Photosynthesis: Active site for water oxidation in PSII
Methane monooxygenase (MMO):
Relative energy of two states (Q (bis-μ-oxo) and P (peroxo))

\[ \Delta E(Q-P) = +3.8 \text{ kcal/mol} \]
\[ \text{lacv3p}^{\text{**}} \text{ basis} \]
\[ \text{Dielectric effect} = -0.7 = +3.1 \text{ kcal/mol} \]

\[ \Delta E(Q-P) = -9.5 \text{ kcal/mol} \]
\[ \text{lacv3p}^{\text{**}} \text{ basis} \]
\[ \text{Dielectric effect} = +3.7 = -5.8 \text{ kcal/mol} \]
Conclusion: Changes has occurred in the backbone, which are not related to the differences between P and Q.
Large model: Take the peptide chain from Q and place it in P:

\[ \Delta(Q-P) = +3.3 \text{ kcal/mol} \]

Compare small model: +3.1 kcal/mol

Conclusion: The large model has no effect on \( \Delta(Q-P) \), it only makes the calculation more difficult.

This kind of problem is present in all large models, even worse in very large QM/MM models.

Solution: Sampling / complete realistic model
QM/MM: Partition the molecular system:

Generic QM/MM energy:

\[ E_{QM/MM} = E_{QM}^{model} + E_{MM}^{(r-m)} + E_{MM}^{(r-m)*model} \]

where the term \( E_{MM}^{(r-m)*model} \) describes certain interactions between the (r-m) and model parts

Gaussian: ONIOM energy: (mechanical embedding)

\[ E_{ONIOM} = E_{QM}^{model} + E_{MM}^{real} - E_{MM}^{model} \]

where \( E_{MM}^{real} - E_{MM}^{model} \) includes the interaction between the (r-m) and the model parts
ONIOM (Gaussian)

**ONIOM**: our Own N-layer Integrated molecular Orbital molecular Mechanics (compare onion)

Any methods can be combined, and the above energy expressions should rather be labelled high level (instead of QM) and low level (instead of MM). Thus different levels of QM methods can be combined.

ONIOM is a **multilayer** method, thus more than two layers can be used with different methods for each layer, e.g. B3LYP/HF/MM
Examples of system partitioning:

Simple case with no bonds across the border line

More complicated case with bonds across the border line

Bonds across the border line are treated e.g. by the addition of link atoms in the model part (ONIOM) or by the use of localized frozen orbitals (Qsite)
Electronic embedding

The above energy expressions treat the interaction between the model part and the (real-model) part at the MM-level, and is therefore referred to as mechanical embedding. This means that there is no polarization effects between the two parts included.

If the electrostatic interactions between the model and the (real-model) parts are treated at the QM level, this is referred to as electronic embedding.

In the formulas below polarization effects from the (real-model) part on the model part are thus included.
\[ E_{\text{ONIOM-EE}} = E_{\text{QM},v}^{\text{model}} + E_{\text{MM}}^{\text{real}} - E_{\text{MM}}^{\text{model}} - \sum_{J} \sum_{N} (q_{J} S_{N} q_{N})/r_{JN} \]

where \( N \) refers to the (real-model) and \( J \) refers to the model part. \( S_{N} \) is a scale factor, zero for atoms close to each other and one for atoms three bonds or more apart.

\( E_{\text{QM},v}^{\text{model}} \) is calculated using a Hamiltonian that includes charges in the (real-model) part (scaled as above). Therefore the Coulomb interaction from the MM-calculations has to be subtracted as above.

Electronic embedding gives more accurate energies and should be used.

Artificial wavefunction localization in real-model can occur
Geometry optimization in QM/MM

Iterative schemes of different kinds where the QM and the MM parts are individually optimized.

Derivatives are constructed for the whole energy expression and the geometry of the whole system is optimized as in a normal QM calculation. This procedure is e.g. used in ONIOM.

Geometry optimization in ONIOM can at present only be done using mechanical embedding. Electronic embedding is determined from single point calculations on the optimized structures.
Relative energies and chemical reactions in QM/MM

Beware of local minima in MM-part, not possible to inspect

Iterate back and forth between different stationary points to be sure to be on the same local minima everywhere

These problems are the same for all large models
Example: Naphthalene 1,2-dioxygenase (NDO)

\[
\begin{align*}
\text{Reactant} & \quad 17.5 \quad \text{TS1} \\
\text{Epoxide} & \quad -25.5 \quad \text{TS2} \\
\text{Cation} & \quad -18.3 \quad 
\text{TS3} \\
\text{Cis-dihydrodiol} & \quad -48.1
\end{align*}
\]
SUBSTRATE: Naphthalene

QM-model of the enzyme

ONIOM-model of the enzyme
NDO results: QM calculations
NDO results: ONIOM calculations

\[ \Delta E (\text{kcal mol}^{-1}) \]

ONIOM model

\[ \text{REAC} \ 0.0 \]

\[ \text{TS} \ 28.0 \]

\[ \text{QM}_{\text{BIG}}: 24.4 \]

\[ \text{EPOXIDE} \ -12.9 \]
Sampling phase space in complex systems

(In simulation methods in statistical physics, FK7029, 7.5 credits. Spring semester)

In systems with many degrees of freedom, it is impossible to scan the whole potential energy landscape.

Monte Carlo Simulations:
Algorithm to sample the distribution corresponding to $V_N(X)$

Molecular Dynamics simulations:
The classical motion of the atoms are calculated:

\[ F_i = m a_i \quad F_i = - \frac{\partial V_N(x_i)}{\partial x_i}. \]

Ab initio MD: Forces calculated on the fly.
Classical MD: Pre-calculated analytical force field.
Multiple Length Scales

- MM part
  - > 1000 solute atoms
  - > 10000 solvent atoms

- Interface region

- QM part
  - ~ 100 atoms
  - ~ 400 electrons

⇒ time evolution of a mixed QM/MM system