Introduction to Molecular Mechanics

C. David Sherrill

School of Chemistry and Biochemistry

Georgia Institute of Technology

Introduction

Molecular Mechanics or *force-field methods* use classical type models to predict the energy of a molecule as a function of its conformation. This allows predictions of

- Equilibrium geometries and transition states
- Relative energies between conformers or between different molecules

Molecular mechanics can be used to supply the potential energy for molecular dynamics computations on large molecules. However, they are not appropriate (nor are most *ab initio* methods!) for bond-breaking reactions.

Stretching Interactions

Express energy due to stretching bonds as a Taylor series about the equilibrium position R_e:

$$E(R) = k_2(R - R_e)^2 + k_3(R - R_e)^3 + \cdots \qquad (1)$$

where R is a bond length. The first term should dominate.

How do we get the *parameters* (k_2 and k_3, etc.)? Could get from experiment or by a full quantum mechanical calculation. *The central idea of molecular mechanics is that these constants are* **transferrable** *to other molecules.* Most C–H bond lengths are 1.06 to 1.10 Å in just about any molecule, with stretching frequencies between 2900 and 3300 cm^{-1}. This strategy is refined using different "atom types."

Figure 1: Atom Types for MM2

Table 2.1 MM2(91) atom types

Type	Symbol	Description	Type	Symbol	Description
1	C	sp^3-carbon	28	H	enol or amide
2	C	sp^2-carbon, alkene	48	H	ammonium
3	C	sp^2-carbon, carbonyl, imine	36	D	deuterium
4	C	sp-carbon	20	lp	lone pair
22	C	cyclopropane	15	S	sulfide (R_2S)
29	C·	radical	16	S+	sulfonium (R_3S^+)
30	C+	carbocation	17	S	sulfoxide (R_2SO)
38	C	sp^2-carbon, cyclopropene	18	S	sulfone (R_2SO_2)
50	C	sp^2-carbon, aromatic	42	S	sp^2-sulfur, thiophene
56	C	sp^3-carbon, cyclobutane	11	F	fluoride
57	C	sp^2-carbon, cyclobutene	12	Cl	chloride
58	C	carbonyl, cyclobutanone	13	Br	bromide
67	C	carbonyl, cyclopropanone	14	I	iodide
68	C	carbonyl, ketene	26	B	boron, trigonal
71	C	ketonium carbon	27	B	boron, tetrahedral
8	N	sp^3-nitrogen	19	Si	silane
9	N	sp^2-nitrogen, amide	25	P	phosphine (R_3P)
10	N	sp-nitrogen	60	P	phosphor, pentavalent
37	N	azo or pyridine (–N=)	51	He	helium
39	N+	sp^3-nitrogen, ammonium (R_4N^+)	52	Ne	neon
40	N	sp^2-nitrogen, pyrrole	53	Ar	argon
43	N	azoxy (–N=N–O)	54	Kr	krypton
45	N	azide, central atom	55	Xe	xenon
46	N	nitro ($-NO_2$)	31	Ge	germanium
72	N	imine, oxime (=N–)	32	Sn	tin
6	O	sp^3-oxygen	33	Pb	lead (R_4Pb)
7	O	sp^2-oxygen, carbonyl	34	Se	selenium
41	O	sp^2-oxygen, furan	35	Te	tellurium
47	O^-	carboxylate	59	Mg	magnesium
49	O	epoxy	61	Fe	iron(II)
69	O	amine oxide	62	Fe	iron(III)
70	O	ketonium oxygen	63	Ni	nickel(II)
5	H	hydrogen, except on N or O	64	Ni	nickel(III)
21	H	alcohol (OH)	65	Co	cobalt (II)
23	H	amine (NH)	66	Co	cobalt (III)
24	H	carboxyl (COOH)			

Note that special atom types are defined for carbon atoms involved in small rings, like cyclopropane and cyclobutane. The reason for this will be discussed in Section 2.2.2.

The Force-Field

Molecular mechanics expresses the total energy as a sum of Taylor series expansions for *stretches* for every pair of bonded atoms, and adds additional potential energy terms coming from *bending*, *torsional energy*, *van der Waals energy*, *electrostatics*, and *cross terms*:

$$E = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}. \qquad (2)$$

By separating out the van der Waals and electrostatic terms, molecular mechanics attempts to make the remaining constants more transferrable among molecules than they would be in a spectroscopic force field.

History

- D. H. Andrews (*Phys. Rev.*, 1930) proposed extending spectroscopic force field ideas to doing molecular mechanics
- F. H. Westheimer (1940) performed the only molecular mechanics calculation done by hand to determine the transition state of a tetrasubstituted biphenyl
- J. B. Hendrickson (1961) performs conformational analysis of larger than 6 membered rings
- K. B. Wiberg (1965) publishes first *general* molecular mechanics type program with ability to find energy minimum
- N. L. Allinger [*Adv. Phys. Org. Chem.* **13**, 1 (1976)]

publishes the first (MM1) in a series of highly popular force fields; the second, MM2, follows in 1977

- Many other force field methods have been developed over the years

Stretch Energy

The stretching potential for a bond between atoms A and B is given by the Taylor series

$$\begin{aligned} E(R^{AB}) &= k_2^{AB}(R^{AB} - R_0^{AB})^2 + k_3^{AB}(R^{AB} - R_0^{AB})^3 \\ &+ k_4^{AB}(R^{AB} - R_0^{AB})^4 + \cdots \end{aligned} \quad (3)$$

and different force field methods retain different numbers of terms in this expansion (frequently, only the first term is kept). Note that such expansions have incorrect limiting behavior at large distances.

Morse Potentials for Stretch Energy

A simple function with correct limiting behavior is the Morse potential

$$E_{\mathrm{str}}(R - R_0) = D[1 - e^{\sqrt{k/2D}(R-R_0)}]^2, \tag{4}$$

where D is the dissociation energy. However, this potential gives very small restoring forces for large R and therefore causes slow convergence in geometry optimization. For this reason, the truncated polynomial expansion is usually preferred.

Figure 2: Stretching Potential for CH_4 with exact, 2nd order, 4th order, and Morse potentials (Jensen, Introduction to Computational Chemistry)

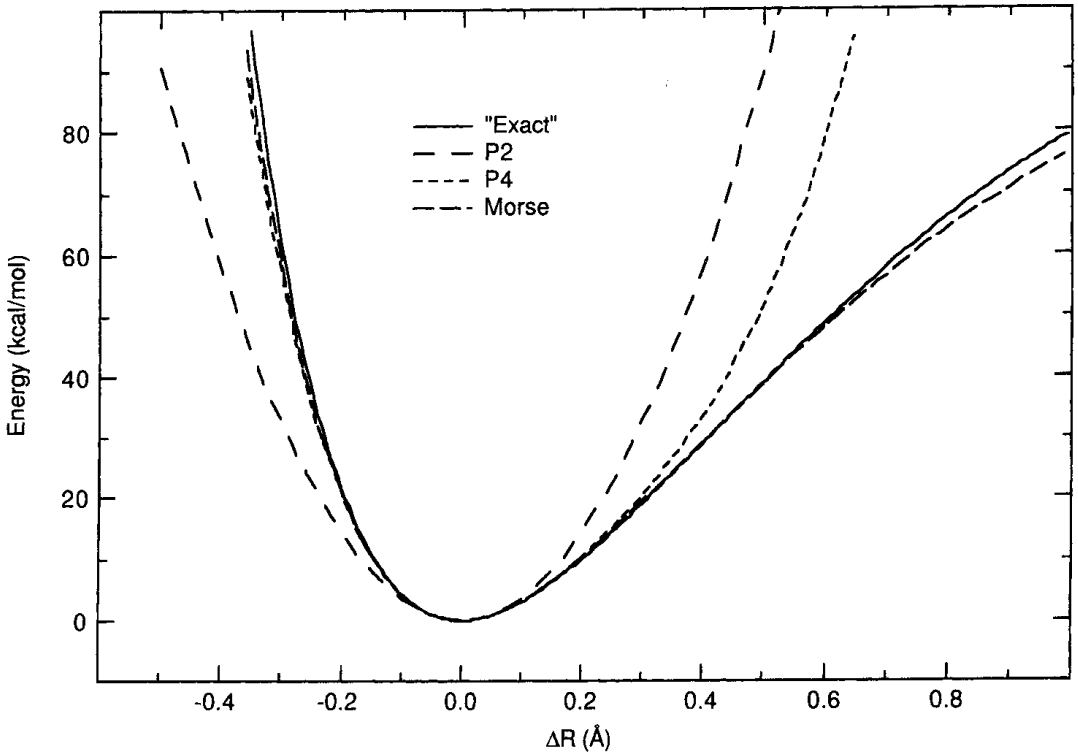

Figure 2.1 The stretch energy for CH_4

Bend Energy

Bending energy potentials are usually treated very similarly to stretching potentials; the energy is assumed to increase quadratically with displacement of the bond angle from equilibrium.

$$E_{\text{bend}}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})^2 \qquad (5)$$

An unusual thing happens for $\theta^{ABC} = 180^{\text{o}}$: the derivative of the potential needs to go to zero. This is sometimes enforced (Fig. 3).

Figure 3: Bending Potential for H_2O (Jensen)

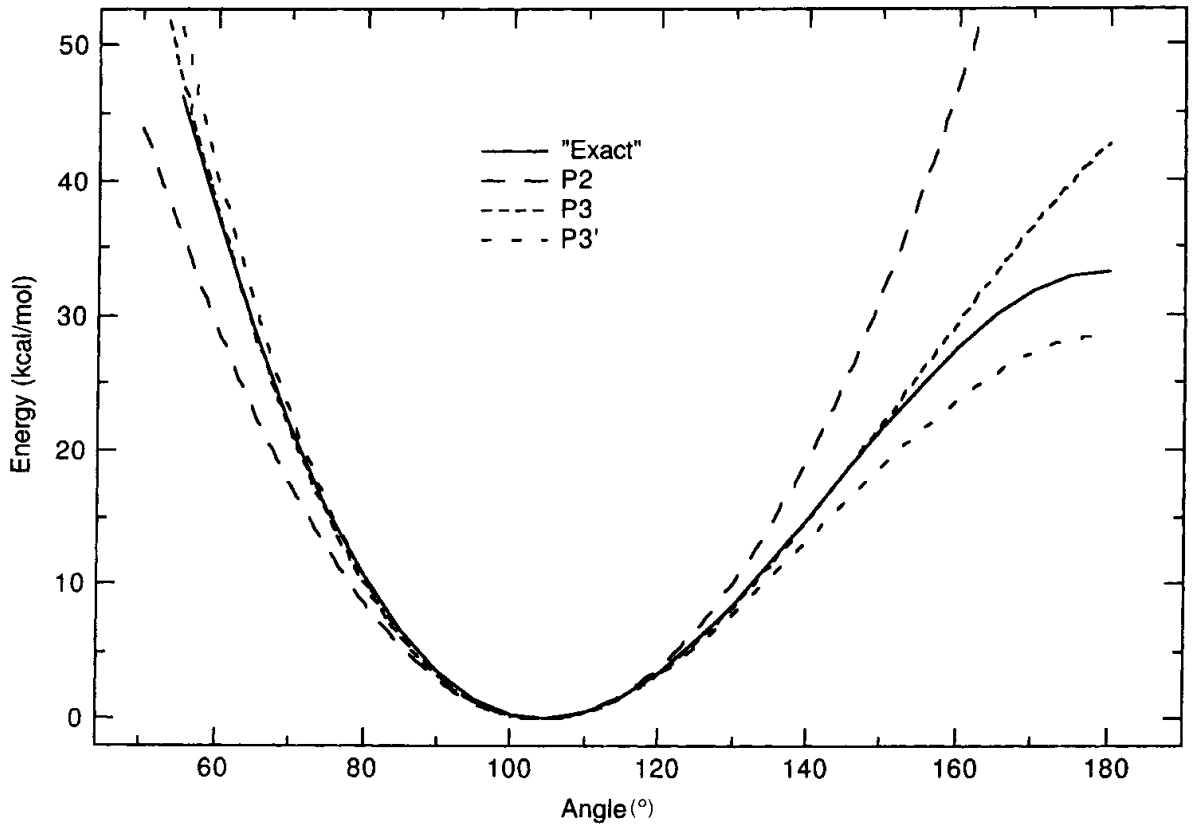

Figure 2.5 The bending energy for H_2O

Out-of-Plane Bending

The potential for moving an atom out of a plane is sometimes treated separately from bending (although it also involves bending). An out-of-plane coordinate (either χ or d) is displayed below. The potential is usually taken quadratic in this out-of-plane bend,

$$E_{\text{bend}-\text{oop}}(\chi^B) = k^B(\chi^B)^2. \tag{6}$$

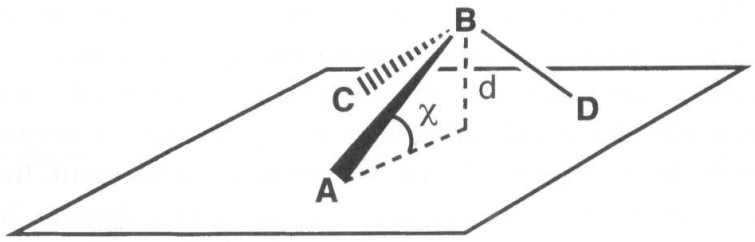

Figure 2.6 Out-of-plane variable definitions

Torsional Energy

The torsional energy term attempts to capture some of the steric and electrostatic nonbonded interactions between two atoms A and D connected through an intermediate bond B–C. The torsional angle ω (also often denoted τ) is depicted below. It is the angle betwen the two planes defined by atoms A, B, and C and by B, C, and D.

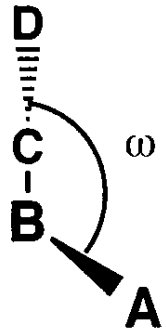

Figure 2.7 Torsional angle definition

Form of the Torsional Potential

The torsional potential is not expanded as a Taylor series because the torsional angle can go far from equilibrium. Fourier series are used instead:

$$E_{\text{tors}}(\omega^{ABCD}) = \sum_{n=1} V_n^{ABCD} cos(n\omega^{ABCD}). \tag{7}$$

Often this is rewritten to make sure the energy is non-negative, and typically the number of terms is 3 (bad for octahedral!).

$$\begin{aligned} E_{\text{tors}}(\omega^{ABCD}) &= \frac{1}{2} V_1^{ABCD}[1 + cos(\omega^{ABCD})] \\ &+ \frac{1}{2} V_2^{ABCD}[1 - cos(2\omega^{ABCD})] \\ &+ \frac{1}{2} V_3^{ABCD}[1 + cos(3\omega^{ABCD})]. \end{aligned} \tag{8}$$

For a molecule like ethylene, rotation about the C=C bond must be periodic by 180°, so only even terms $n = 2, 4, \ldots$ can occur. For a molecule like ethane, only terms $n = 3, 6, 9, \ldots$ can occur.

Figure 4: Torsional Potentials for $n = 3$ and 2 (Jensen)

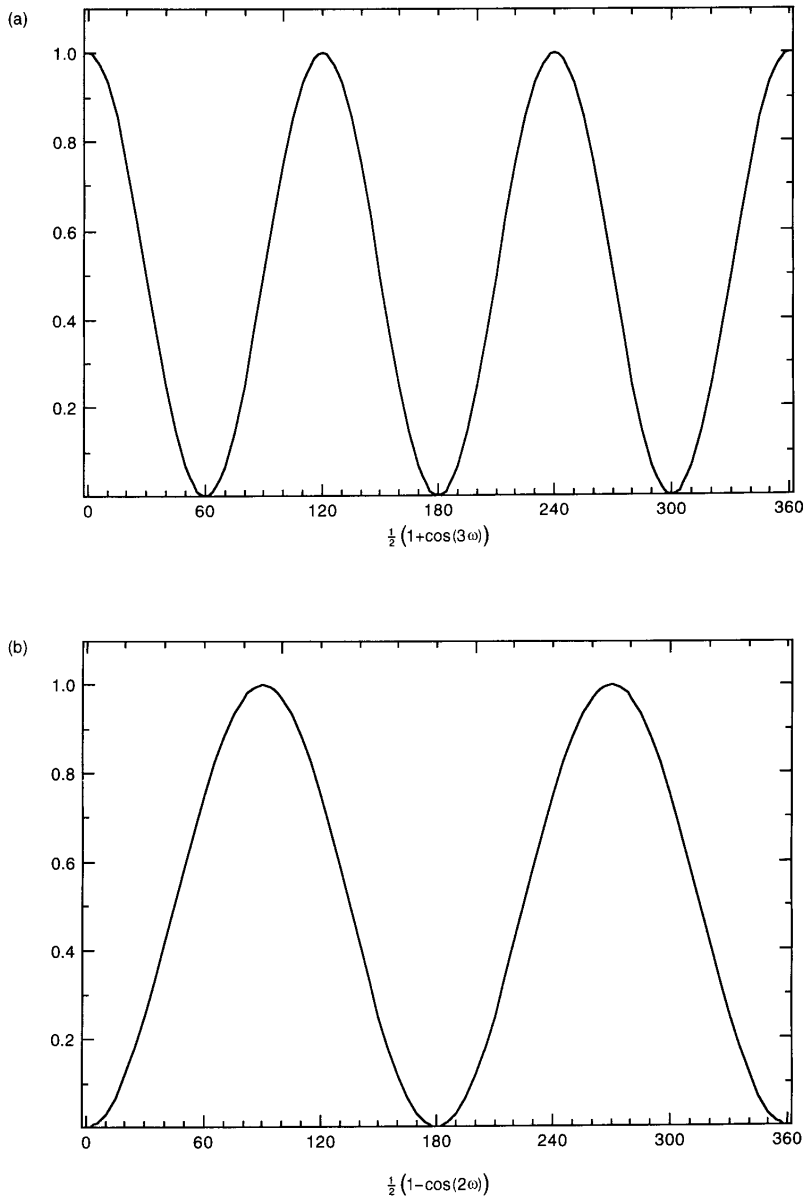

Figure 2.8 Torsional energy functions

Figure 5: Example Torsional Potential for $n = 1$ (Jensen)

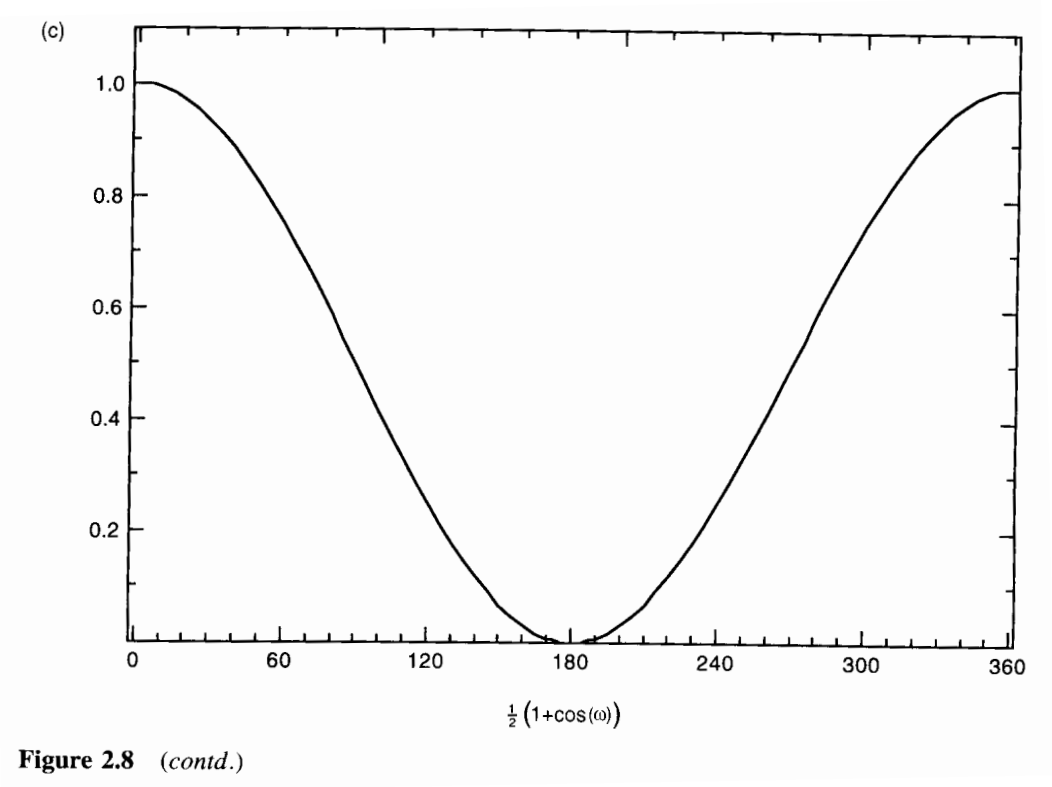

Figure 2.8 (*contd.*)

van der Waals Energy

The van der Waals energy arises from the interactions between electron clouds around two nonbonded atoms.

Short range: strongly repulsive

Intermediate range: attractive

Long range: goes to zero

The attraction is due to *electron correlation* which results in "dispersion" or "London" forces (instantaneous multipole / induced multipole).

Form of the van der Waals Energy

van der Waals energies are usually computed for atoms which are connected by no less than two atoms (e.g., 1-4 interactions between A and D in A-B-C-D and higher). Interactions betwen atoms closer than this are already accounted for by stretching and/or bending terms.

At intermediate to long ranges, the attraction is proportional to $1/R^6$. At short ranges, the repulsion is close to exponential. Hence, an appropriate model of the van der Waals interaction is

$$E_{\text{vdw}}(R^{AB}) = Ce^{-DR} - \frac{E}{R^6}. \tag{9}$$

Since the van der Waals and electrostatic interactions are long-range, they can become the dominant costs of force-field computation. The van der Waals term can be speeded up substantially by a more economical expression, the Lennard-Jones potential

$$E_{\text{vdw}}(R^{AB}) = \epsilon \left[\left(\frac{R_0}{R} \right)^{12} - 2 \left(\frac{R_0}{R} \right)^{6} \right]. \tag{10}$$

The R^{-12} term is easier to compute than the exponential because no square roots need to be taken to get R. Figure 6 presents a comparison of various van der Waals potentials.

Figure 6: Example van der Waals Potentials (Jensen)

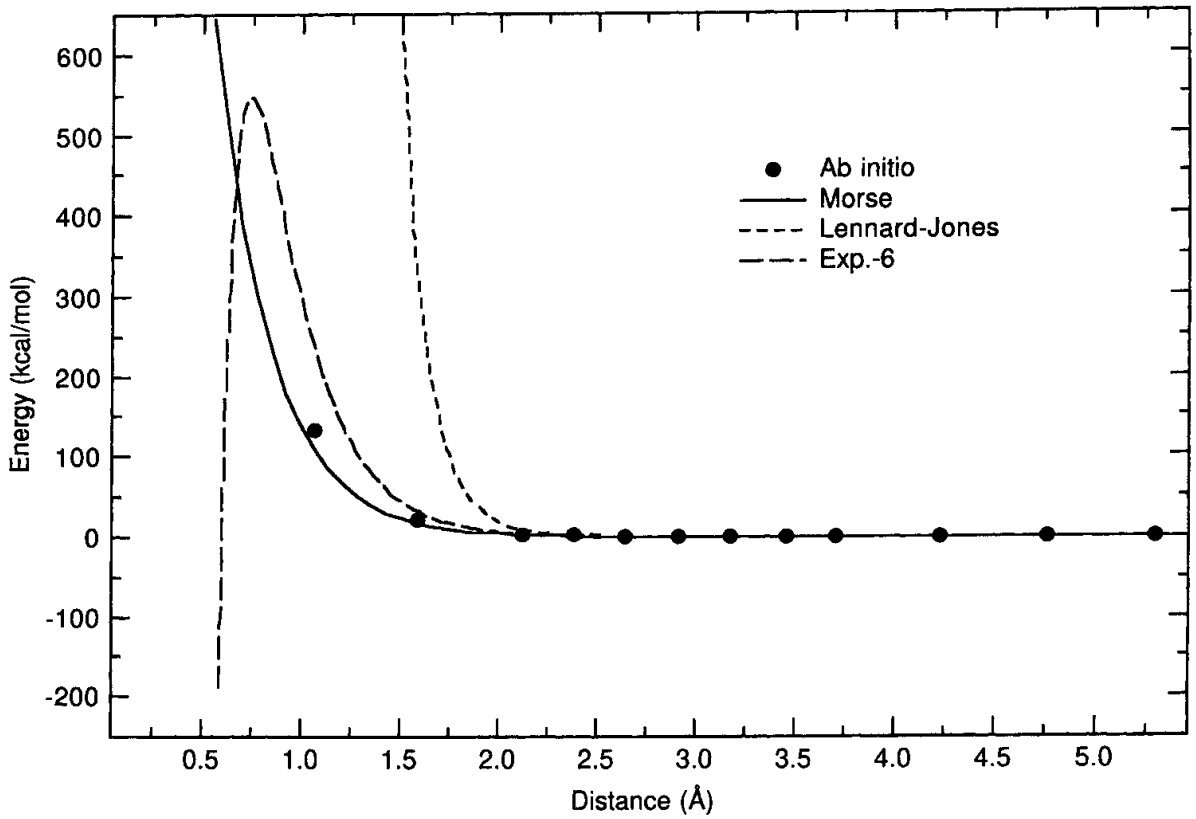

Figure 2.9 Comparison of E_{vdw} functionals for the H_2–He potential

Electrostatic Energy

Electrostatic terms describe the Coulomb interaction between atoms A and B with partial charges (e.g., in carbonyls), according to

$$E_{\mathrm{el}}(R^{AB}) = \frac{Q^A Q^B}{\epsilon R^{AB}}, \tag{11}$$

where ϵ is an effective dialectric constant which is 1 in vacuum but higher when there are intermediate atoms or solvent. Usually ϵ is picked fairly arbitrarily; higher values or so-called "distance-dependent dialectrics" ($\epsilon = \epsilon_0 R^{AB}$) account for "screening" and kill off the electrostatic contributions faster, making them easier to compute.

Alternative Approaches to Electrostatics

A slightly different approach to partial charges is to consider polarized bonds as dipoles, and compute the electrostatic interaction between these dipoles (e.g., MM2 and MM3):

$$E_{\rm el}(R^{AB}) = \frac{\mu^A \mu^B}{\epsilon (R^{AB})^3} \left(cos\chi - 3 \cos\alpha_A cos\alpha_B \right), \qquad (12)$$

where χ is the angle between the dipoles and α_A and α_B are the angles each dipole makes with the line joining atoms A and B.

Fitting Atomic Charges

Atoms don't have charges!! Quantum mechanics tells us electrons are smeared out, and their charge is shared among nearby atoms. No unique way to assign electrons to atoms.

Nevertheless, "atomic charges" are a useful approximation. Most favored procedure is to perform a quantum computation and then choose charges that reproduce the quantum electrostatic potential. Restrained Electrostatic Potential (RESP) procedure used for AMBER force field.

$$\phi_{esp}(\mathbf{r}) = \sum_{A}^{nuc} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\Psi^*(\mathbf{r}')\Psi(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \qquad (13)$$

Computational Cost of Electrostatic Terms

Like van der Waals terms, electrostatic terms are typically computed for nonbonded atoms in a 1,4 relationship or further apart. They are also long range interactions and can dominate the computation time.

The number of nonbonded interactions grows *quadratically* with molecule size. The computation time can be reduced by cutting off the interactions after a certain distance. The van der Waals terms die off relatively quickly ($\propto R^{-6}$) and can be cut off around 10 Å. The electrostatic terms die off slower ($\propto R^{-1}$, although sometimes faster in practice), and are much harder to treat with cutoffs. This is a problem of interest to developers.

Improved Electrostatic Energies

The point-charge model has serious deficiencies: (a) electrostatic potentials are not accurately reproduced; (b) simple models don't allow the charges to change as the molecular geometry changes, but they should; (c) only pairwise interactions are considered, but an electrostatic interaction can actually change by ∼10-20% in the presence of a third body (induction or "polarization" effects).

Improved Electrostatic Potentials

Electrostatic potentials can be more accurately reproduced by:

- Allowing non-atom-centered charges
- Adding point dipoles, quadrupoles, etc.

Both are employed in Anthony Stone's *Distributed Multipole Analysis (DMA)*. However, this does not yet allow the electrostatic variables to change as a function of geometry or to respond to electrostatic potentials generated by nearby atoms. This requires *polarizable force-fields* such as the fluctuating charge model.

The Fluctuating Charge Model

Consider the energy as a function of the number of electrons N:

$$E = E_0 + \frac{\partial E}{\partial N}\Delta N + \frac{1}{2}\frac{\partial^2 E}{\partial N^2}(\Delta N)^2 + \cdots \tag{14}$$

Electronetativity $\chi = -\partial E/\partial N$. Chemical hardness $\eta = \partial^2 E/\partial N^2$. Setting $\Delta Q = -\Delta N$ and expanding about the point with no net atomic charges,

$$E = \chi Q + \frac{1}{2}\eta Q^2 + \cdots \tag{15}$$

Considering the electrostatic energy also depends on the external potential ϕ, and summing over all sites (usually

atomic centers),

$$E_{el} = \sum_A \phi_A Q_A + \sum_A \chi_A Q_A + \frac{1}{2}\eta_{AB} Q_A Q_B. \qquad (16)$$

In vector notation, and requiring that the energy be stationary with respect to the charges with and without an external potential,

$$\frac{\partial E}{\partial \mathbf{Q}}|_{\phi \neq 0} = \phi + \chi + \eta Q = 0 \qquad (17)$$

$$\frac{\partial E}{\partial \mathbf{Q}}|_{\phi = 0} = \chi + \eta Q^0 = 0 \qquad (18)$$

which yields

$$\Delta \mathbf{Q} = -\eta^{-1}\phi, \qquad (19)$$

where ϕ depends on the charges at all sites (and thus the

equations must be solved iteratively),

$$\phi(\mathbf{r}) = \sum_A \frac{Q_A}{|\mathbf{r} - \mathbf{R}_A|}. \tag{20}$$

Other Polarizable Models

The fluctuating charge model has some limitations. More general models incorporate explicit polarizability terms. A dipole moment can result from a polarization of the electronic charge as a result of an electric field ($\mathbf{F} = \partial\phi/\partial\mathbf{r}$):

$$\mu_{ind} = \alpha F \tag{21}$$

$$E_{el}^{pol} = \frac{1}{2}\mu_{ind}\mathbf{F} = \frac{1}{2}\alpha\mathbf{F}^2 \tag{22}$$

It is common to break up the polarizability α into "atomic polarizability" contributions.

Cross Terms

Cross terms are required to account for some interactions affecting others. For example, a strongly bent H_2O molecule tends to stretch its O–H bonds. This can be modeled by cross terms such as

$$E_{\text{str/bend}} = k^{ABC}(\theta^{ABC} - \theta_0^{ABC}) \left[(R^{AB} - R_0^{AB}) + (R^{BC} - R_0^{BC}) \right] \tag{23}$$

Other cross terms might include stretch-stretch, bend-bend, stretch-torsion, bend-torsion, etc. Force field models vary in what types of cross terms they use.

Parameterizing the Force Fields

Molecular mechanics requires many parameters, e.g., R_0^{AB}, k^{AB}, θ_0^{ABC}, k^{ABC}, V_n^{ABCD}, etc. Assuming there are 30 atoms which form bonds with each other, there are $30^4/2$ torsional parameters for each V_n^{ABCD} term, or 1 215 000 parameters for V_1 through V_3! Only the 2466 "most useful" torsional parameters are present in MM2, meaning that certain torsions cannot be described.

Lack of parameters is a serious drawback of all force field methods. Automatic guessing of unavailable parameters is sometimes done but is very dangerous.

Heats of Formation

ΔH_f is the heat conent relative to the elements at standard state at 25° C (g). This is a useful quantity for comparing the energies of two conformers of a molecule or two different molecules.

Bond energy schemes estimate the overall ΔH_f by adding tabulated contributions from each type of bond. This works acceptably well for strainless systems.

Molecular mechanics adds steric energy to the bond/structure increments to obtain better estimates of ΔH_f.

The bare molecular mechanics energy is not a meaningful quantity for comparing different molecules.

Accuracy of Heats of Formation

In principle, other corrections should be added (but usually are not): population increments (for low-lying conformers), torsional increments (for shallow wells), and corrections for low (< 7 kcal/mol) barriers other than methyl rotation (already included in group increment).

The performance of the MM2 force field for typical molecules is given in Table 1. Overall, these results are rather good; however, unusual molecules can exhibit far larger errors.

Table 1: Average errors in heat of formations (kcal/mol) by MM2[a]

Compound type	Avg error ΔH_f
Hydrocarbons	0.42
Ethers and alcohols	0.50
Carbonyls	0.81
Aliphatic amines	0.46
Aromatic amines	2.90
Silanes	1.08

[a]Table 2.6 from Jensen's *Introduction to Computational Chemistry*.

Different Force Field Methods

Class I Methods: Higher order terms and cross terms. Higher accuracy, used for small or medium sized molecules. Examples: Allinger's MM1-4, EFF, and CFF.

Class II Methods: For very large molecules (e.g., proteins). Made cheaper by using only quadratic Taylor expansions and neglecting cross terms. Examples: AMBER, CHARMM, GROMOS, etc. Made even cheaper by using "united atoms."

Hybrid Force Field/Electronic Structure Methods

Treat "uninteresting" parts of the molecule by force field methods and "interesting" parts by high-accuracy electronic structure methods. This approach is useful for systems where part of the molecule is needed at high accuracy or for which no force field parameters exist (e.g., metal centers in metalloenzymes).

Examples: Morokuma's ONIOM method; QM/MM methods.

QM/MM Approaches

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \tag{24}$$

- Mechanical embedding: Include bonded and steric interactions between QM and MM regions, and maybe assign partial charges to the QM atoms
- Electronic embedding: Also add partial charges on MM atoms to QM computation
- Polarizable embedding: Also allow QM atoms to polarize MM region (only applies to polarizable force fields or Gordon's effective fragment method)

If the QM/MM boundary cuts covalent bonds, dangling bonds are terminated by so-called "link atoms" (typically hydrogen).

Morokuma's ONIOM Approach

ONIOM: Our Own n-Layered Integrated Molecular Orbital Molecular Mechanics Method

$$
\begin{aligned}
E_{ONIOM}(\text{real system, high level}) &= E_{low\ level}(\text{real system}) \\
&+ E_{high\ level}(\text{model system}) \\
&- E_{low\ level}(\text{model system})
\end{aligned}
$$

Can also involve multiple layers. Originally used only mechanical embedding for the QM/MM interface, but more recently uses electronic embedding. Can also do as a QM/QM with high and low levels of quantum mechanics.