

# Computational Chemistry - part I: Force field methods

Tomas Öberg  
University of Kalmar, Department  
of Biology and Environmental  
Science

*"All models are wrong but some are useful"*

George Box

2003-01-31

# Content of the three seminars

- Today
  - Some concepts in molecular modelling
  - A brief introduction to *ab initio* modelling
  - Empirical force field models: Molecular mechanics
- Next meeting
  - Energy minimisation
  - Optimisation methods
- Last Meeting
  - Chemoinformatics
  - Molecular descriptors
  - QSAR/QSPR, chemometrics

2003-01-31

## Literature - textbooks

- Andrew R. Leach (2001): *Molecular modelling - principles and applications*, 2nd edition, Prentice-Hall, 744 pp.
- Mati Karelson (2000): *Molecular descriptors in QSAR/QSPR*, Wiley, 430 pp.
- Walters, F. H. et al (1991): *Sequential simplex optimization*, CRC Press, 325 pp.
- Harald Martens and Tormod Næs (1989): *Multivariate calibration*, Wiley, 419 pp.

## Some of my own publications related to this lecture series

- Öberg, T. *Importance of the first design matrix in experimental simplex optimization*, Chemometrics and Intelligent Laboratory Systems **44**, 147-151, 1998
- Öberg, T.; Deming, S. N. *Find optimum operating conditions fast*, Chemical Engineering Progress **96**, 53-59, 2000.
- Öberg, T. *Prediction of physical properties for PCB congeners from molecular descriptors*, Internet Journal of Chemistry **4**, U4-U17, 2001.
- Öberg, T. *Prediction of vapour pressures for halogenated diphenyl ether congeners from molecular descriptors*, Environmental Science and Pollution Research **9**, 405-411, 2002.
- Öberg, T. *Prediction of gas chromatographic separation for PBDE congeners from molecular descriptors*, ESS Bulletin, submitted.

2003-01-31

# Basic concepts

- What is molecular modelling?
  - Simplified or idealised description of a molecule
  - Ways to mimic the behaviour of molecules and molecular systems
  - Computational techniques to build or calculate molecule representations
- All these models are concerned with the three-dimensional representation of the structures of molecules

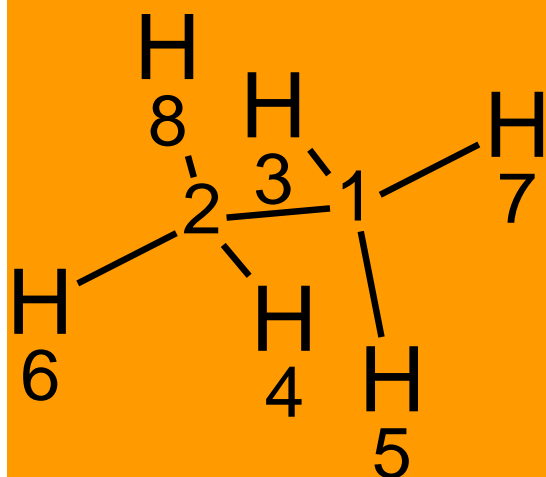
# “Semantics”

- Theoretical chemistry?
  - Quantum mechanics
- Computational chemistry?
  - Quantum mechanics + molecular mechanics, optimisation, simulations, other computer-based methods for understanding and predicting the behaviour of molecular systems (Leach, 2001)

# 3D-representation

- What do we need to represent a molecule and make computer calculations?
  - A coordinate system!
- What kind of coordinate system?
  - Cartesian in three dimensions (x, y, z)
  - Z-matrix, internal coordinates
  - It is possible to convert between the two

## Z-matrix - example ethane (staggered conformation)



1	C						
2	C	1.54	1				
3	H	1.0	1	109.5	2		
4	H	1.0	2	109.5	1	180.0	3
5	H	1.0	1	109.5	2	60.0	4
6	H	1.0	2	109.5	1	-60.0	5
7	H	1.0	1	109.5	2	180.0	6
8	H	1.0	2	109.5	1	60.0	7

Atom, distance (Å), angle and torsion. Internal coordinates (Z-matrix) are commonly used by quantum mechanics software.

# Software

- Numerous computer programs, both for PC and Unix machines.
- Research software as well as commercial, e.g.
  - Hyperchem (PC)
  - Insight (SGI)
- Force fields
  - MM2/MM3/MM4 by Allinger et al
  - AMBER, OPLS, etc.

2003-01-31

# The quantum mechanics approach - why?

- Represents the electrons in calculations. Consequence?
- Possibility to derive properties that depend on the electronic distribution, e.g.
  - Chemical reactions (formation and breakage of bonds)
  - Thermodynamic properties

# A starting point - Schrödinger's equation

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V \right\} \Psi(r) = E \Psi(r)$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V$$

$$H \Psi = E \Psi$$



# Solving the Schrödinger equation

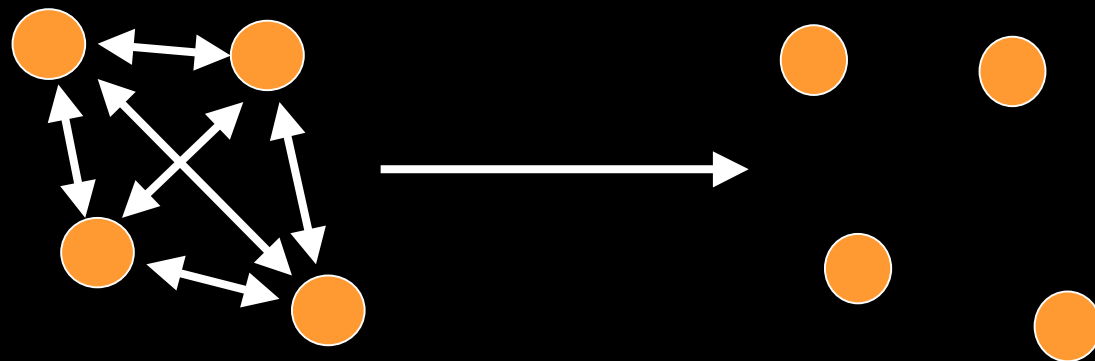
- How to solve this?
  - Find eigenvalues of  $E$  (energy) that correspond to wave functions (eigenfunctions)
  - The energy can be determined by integration over all space
- “Easier said than done”
  - The Schrödinger equation can be solved for a few problems, e.g. a particle in a box, the harmonic oscillator and the hydrogen atom

# The Born-Oppenheimer approximation

- The Schrödinger equation cannot be solved exactly for any molecular system
- Approximations are needed, the first is to decouple the motion of the nuclei from electrons
  - The difference in mass between electrons and protons is 1836
  - $\Psi_{\text{tot}} = \Psi_{\text{electrons}} \Psi_{\text{nuclei}}$

# Hartree theory

- Concludes that the probability of finding an electron at a particular point in space is independent of finding any other electron at that point in space



- Removes the insoluble many-body problem

# The variation theorem

- In a molecular orbital calculation there is no correct solution
- How do we decide if one wave function is better than another?
  - The energy calculated from an approximation of the true wave function will always be greater than the “true” energy
  - Thus, the best wave function is obtained when the energy is a minimum

# *Ab initio* methods

- *Ab initio* is latin for “from the beginning”, so this would imply calculations directly from physical constants, but is it so?
  - Some approximations have already been mentioned, and “calibration calculations” are also needed
  - The basis sets restrict the electrons considered to specific regions in space. These atomic functions are simplified by replacing Slater type orbitals with linear combinations of Gaussian functions
  - A minimal basis set is a representation that contains only the number of functions needed to accommodate all the filled orbitals
  - “*The construction of a new basis set is very much an art*” (Leach, 2001)

# Linear combination of atomic orbitals (LCAO)

- Gaussians to approximate atomic orbitals

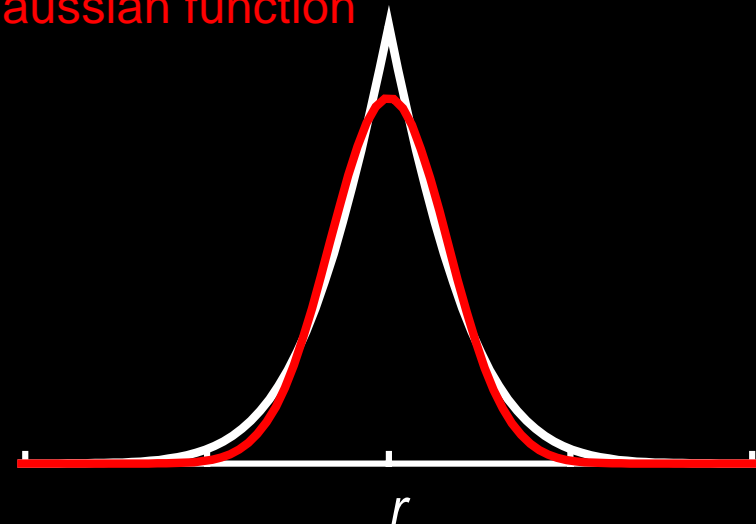
then

- LCAO:

$$\psi = \sum c_r \phi_r$$

Atomic radial function

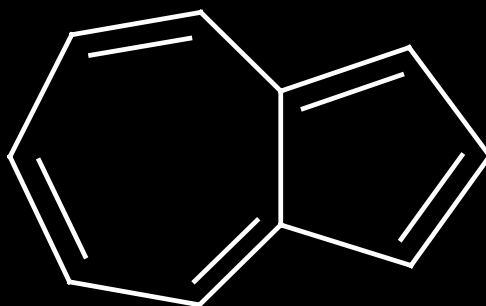
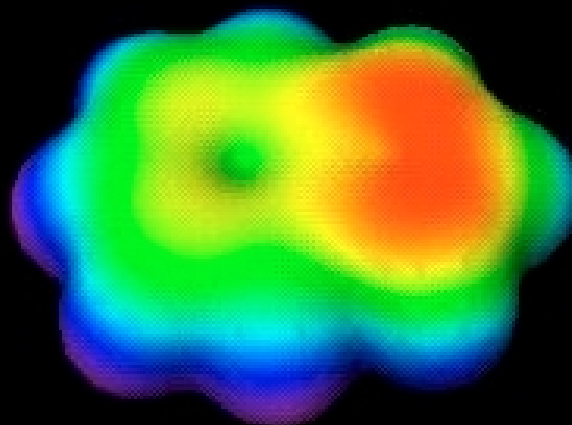
Gaussian function



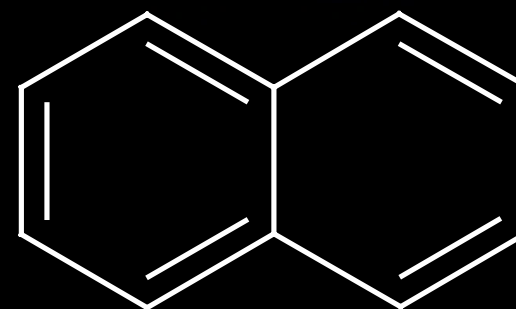
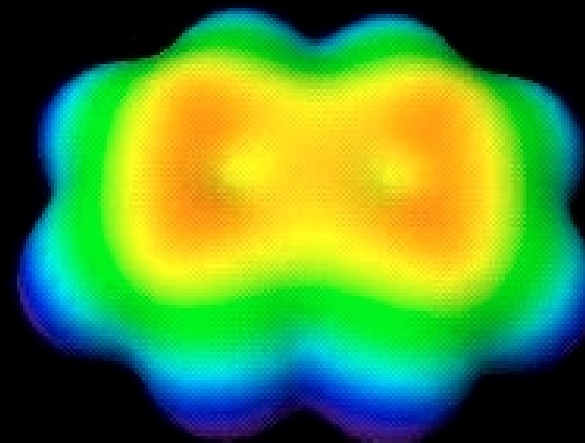
## What can *ab initio* methods accomplish/calculate?

- The dipole moment and other electric multipoles
- Electron density distributions, e.g. highest occupied molecular orbital (HOMO) and lowest unoccupied orbital (LUMO)
- Thermodynamic and structural properties, e.g. rotation barriers

# Electron density distributions



Azulene



Naphthalene

2003-01-31

## Drawbacks with the *ab initio* approach

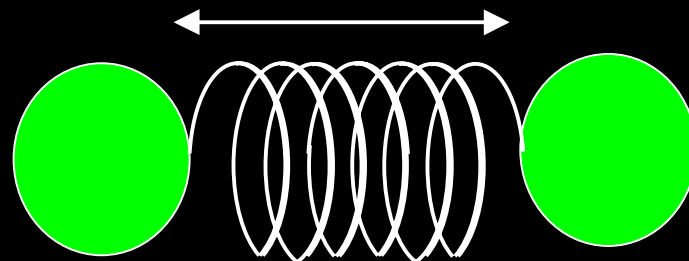
- Time consuming, requires a lot of computational power
- What consequence does this have?
  - Only feasible for small molecules
- Semi-empirical methods is an alternative, e.g. AM1 and PM3
  - Focus only on the valence shell electrons

# The purely empirical approach - molecular mechanics

- What do we need to define molecular geometry (remember the Z-matrix)?
  - Atom types
  - Bond lengths
  - Bond angles
  - Torsional angles (rotation)
- Something more?
  - Interaction between non-bonded parts

## Springs and balls

- Based on classical physics, i.e. atoms are treated as balls and bonds are springs
- The foundation is that molecules are composed of units that are structurally similar



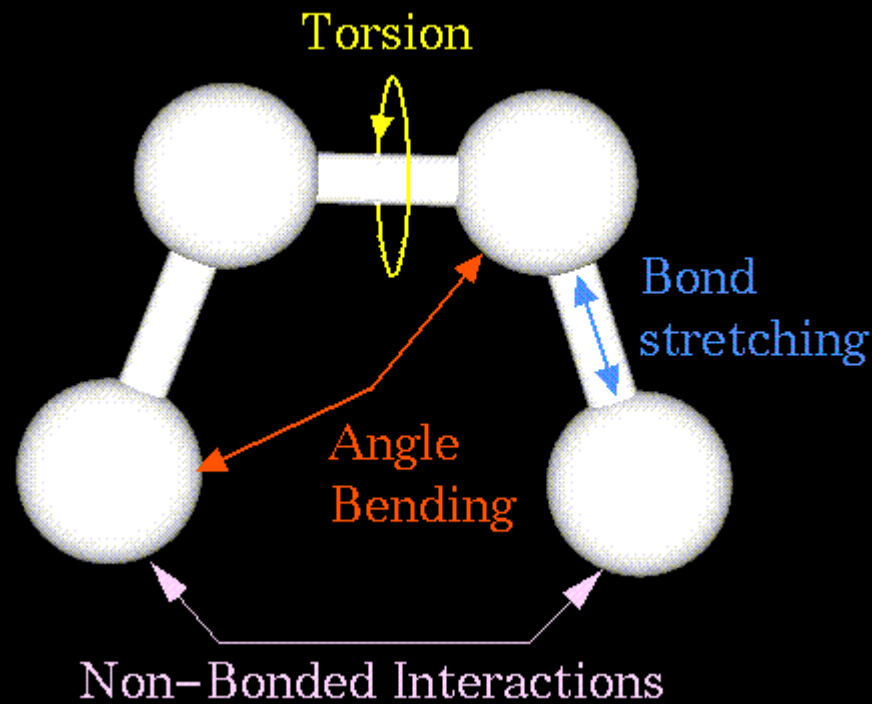
# Limitations - possibilities

- Force field methods (molecular mechanics) ignore electronic motions and calculate only nuclear positions
  - Can this approach be used to predict e.g. the UV-VIS spectra?
  - No, since these are properties that depend upon the electronic distribution
- Can we assume something positive from this limitation?
  - Yes, we can handle much larger systems than with the quantum mechanical calculations

# Assumptions

- Which assumption must be valid to ignore the electrons?
  - The Born-Oppenheimer approximation
- Why must the models be simple?
  - Otherwise no benefit as compared to *ab initio* approach
  - Difficult to parameterise if complex
- Other important attributes?
  - Transferability, something developed for a small number of cases must be possible to apply to a wide range of problems

# Functional form - contributions to the force field



$$V(r^N) = \sum_{bonds} v(l) + \sum_{angles} v(\theta) + \sum_{torsions} v(\omega) + \sum_{non-bonded}$$

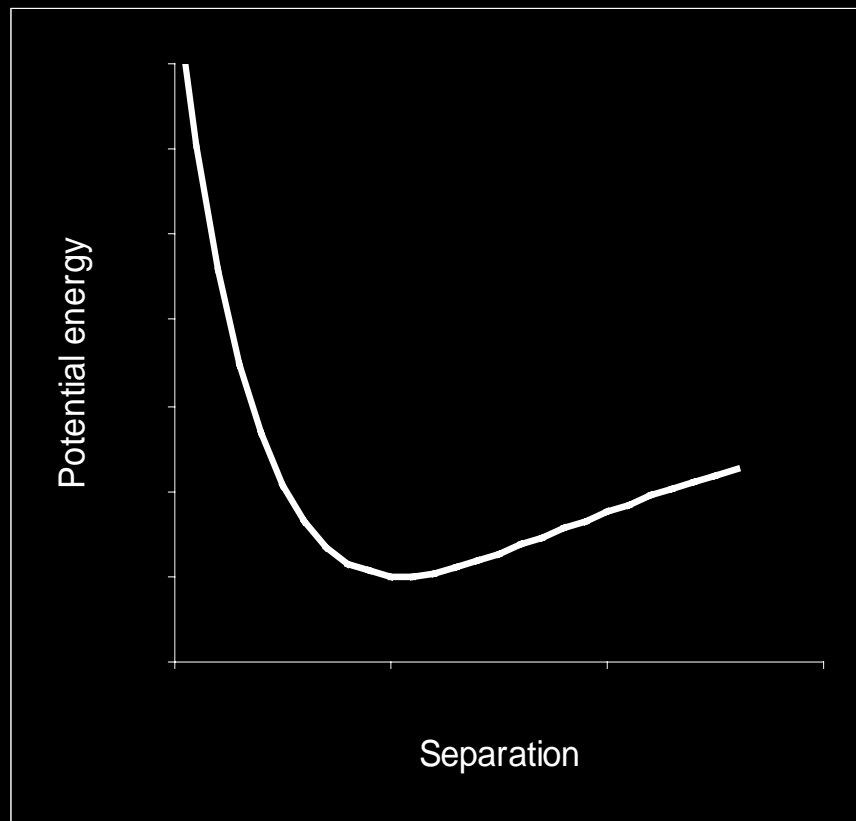
2003-01-31

# Bond stretching



- How does the potential energy change for a typical bond?
  - A minimum at some distance
- Is the change steeper on either side?
  - Yes, the energy increase dramatically when the distance between the nuclei decrease below the minimum

# Bond stretching cont.

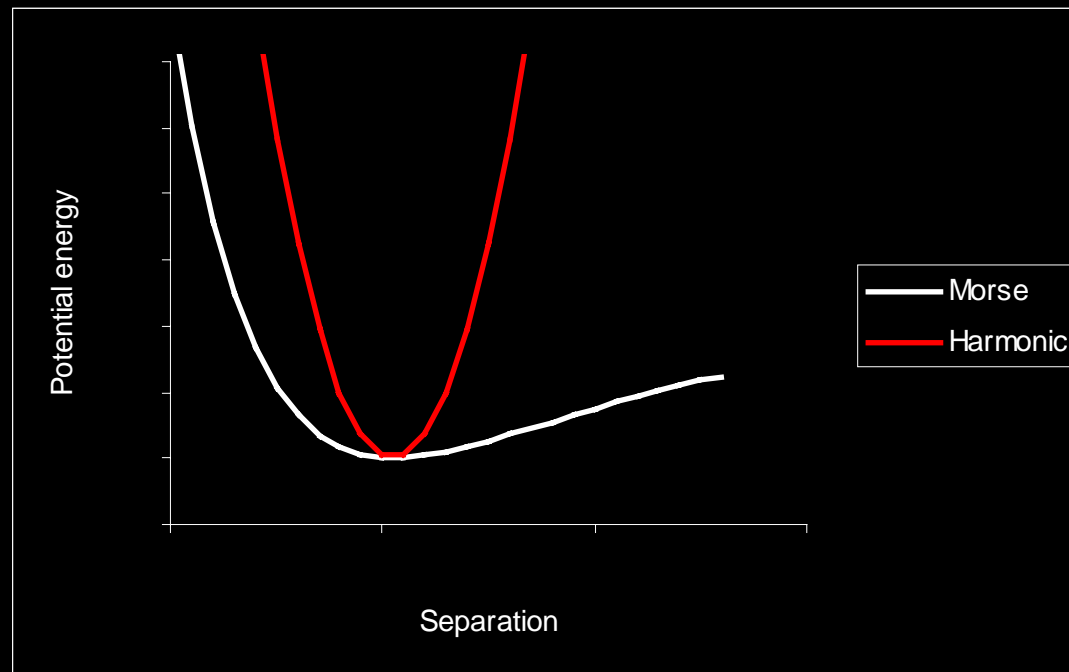


2003-01-31

## Bond stretching cont.

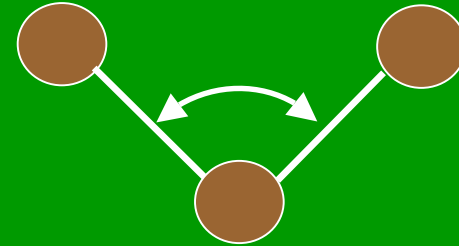
- How can we model this curve?
  - Morse potential  $v(l) = D_e\{1 - \exp[-a(l - l_0)]\}^2$ , where  $l_0$  is the reference bond length
  - Any problem? How many parameters to specify for each bond?
- Can we simplify?
  - A harmonic potential where energy varies with square of distance, e.g.  
$$v(l) = k/2 * (l - l_0)^2$$

# Bond stretching cont.



2003-01-31

# Angle bending



- Can we use a similar approach to model bond angles?
  - Yes, also here a harmonic potential can be applied  $v(\Theta) = k/2 * (\Theta - \Theta_0)^2$
- Is more or less energy needed to distort a bond angle as compared to the bond length?
  - The force constants (k) are much smaller for angle bending

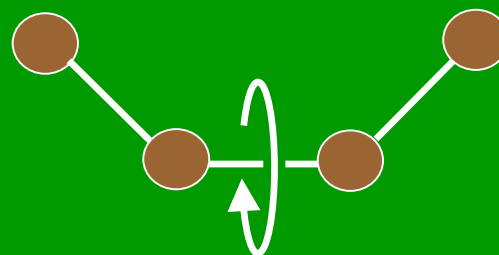
# Improved accuracy

- How can the accuracy of the force field be improved?
  - Adding higher-order terms, i.e. a Taylor-series expansion
  - Bond-stretching, e.g.  $v(l) = k/2 * (l-l_0)^2 [1-k'(l-l_0)-k''(l-l_0)^2 -k'''(l-l_0)^3 \dots]$
  - Angle-bending, e.g.  $v(\Theta) = k/2 * (\Theta-\Theta_0)^2 [1-k'(\Theta-\Theta_0)-k''(\Theta-\Theta_0)^2 -k'''(\Theta-\Theta_0)^3 \dots]$

# Variation in structure and relative energies

- Bond lengths and angles specify the 'hard' degrees of freedom, but are there other contributions to variations in structure and relative energies?
  - The complex interplay between torsional terms and non-bonded interactions
  - The existence of barriers of rotation is a fundamental concept in understanding conformational changes

# Torsional terms



- How can we then represent the changes in torsional potential?
  - Most force fields for organic molecules take each bonded quartet of atoms into consideration
  - Torsional potentials are mostly expressed with a cosine series expansion
  - Improper torsion angles ('out of plane') are incorporated as separate harmonic potentials

## Torsional potential - functional form

$$v(\omega) = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)]$$

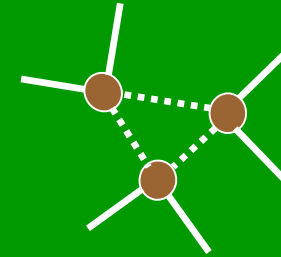
- $\Omega$  is the torsion angle
- $V_n$  a constant related to the barrier “height”
- $n$  is the “multiplicity”, that is the number of minimum points in a 360 rotation
- and  $\gamma$  is the phase factor (determines where the torsion angle pass the minimum)

# Coupling between internal coordinates

- Empirically it is often found that if the bond angle is decreased the neighbouring bonds are stretched to reduce interaction between the 1,3-atoms.
- How can this be modelled?
  - Such interactions can be described with cross terms, e.g.

$$v(l_1, l_2, \theta) = \frac{k_{l_1, l_2, \theta}}{2} [(l_1 - l_{1,0}) + (l_2 - l_{2,0})](\theta - \theta_0)$$

# Non-bonded interactions



- Usually considered in two groups
  - Electrostatic interactions
  - van der Waals interactions
- One aspect is important in modelling both of these, which?
  - Distance in space, usually modelled as a function of the inverse power of distance

# Electrostatic interactions

- Differences in electronegativity between elements give rise to unequal charge distributions within molecules
- The charge distribution can be represented in various ways, some suggestions?
  - Fractional point charges within the molecule
  - Treat the molecule as a single entity and calculate the electric moments (multipoles)
  - Distributed multipoles and point charges
  - Assign dipoles to the bonds in the molecule

# Point charges

- Where do we place the point charges?
  - Nuclear centres (partial atomic charges or net atomic charges)
  - Accurate representation often requires charges to be placed also at locations outside of the nuclei
- How do we then calculate the electrostatic interaction?
  - As the sum of interactions between all pairs of point charges (Coulomb's law)

## Point charges cont.

- Electrostatic interaction between two molecules or different part of the same molecule:

$$V = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

- $N_A$  and  $N_B$  are the number of point charges

## Point charges cont.

- The partial atomic charge is unfortunately not possible to determine, either by observation or from the wave function.
- Can we think of some experiment to assist?
  - For small molecules where the geometry is known, we can calculate the charges to reproduce the electric moments, e.g. dipole
- Some methods calculate atomic charges only from information about atoms present (electronegativity)

# van der Waals interactions

- All non-bonded interactions cannot be accounted for by electrostatic forces
  - van der Waals discovered the deviations from ideal gas behaviour, thus the name
  - London explained the attractive dispersive forces, thus also named “London force”
- Fluctuations in the electron clouds form instantaneous dipoles, consequence?
  - Dipoles are formed in neighbouring atoms, giving rise to an attractive inductive effect

## van der Waals interactions, cont.

- Also a short range repulsive contribution, referred to as exchange forces
  - This effect is due to electrostatic repulsion between pairs of electrons (the Pauli principle)
- A simple empirical expression, *Lennard-Jones 12-6 function*
  - Two adjustable parameters, collision diameter  $\sigma$  and well-depth  $\varepsilon$

$$v(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]$$

# Large molecules

- The number of non-bonded interactions scales with the square of the number of interaction sites present
- This will obviously cause problems when modelling large molecules, e.g. proteins
- Possible solution?
  - Reduce the number of atoms, e.g. subsume hydrogen atoms into the atoms they are bonded to (“united atom force fields”)
- Possible drawback?
  - Chiral centres may be able to invert

## Large molecules, cont.

- Is it possible to simplify even more?
  - An entire group of atoms can be treated as a point of interaction, e.g. a benzene ring
- Simplify more than that?
  - Yes, we can drop the direct connection to the molecule. Often practised in the polymer field, e.g. lattice models or a freely rotating chain model

# Force field parametrisation

- What is the parametrisation?
  - It is the determination of all the parameters used to describe and model the various atom types, bonds, angles, etc. In other words all the constants used in the empirical functions adapted
- How do we select the parameter values?
  - Experimental data, e.g. X-ray methods for bond lengths
  - Quantum mechanics calculations, e.g. torsional barriers
  - Least square fitting for optimal model structure

## Some limitations to molecular mechanics

- The calculations do not account for electrons
- The selection of “atom type” is crucial to the computational result
- Only ground states considered

# Examples of MM2 atom types (>60)

1	C	sp <sup>3</sup> carbon	13	Br	bromine
2	C	sp <sup>2</sup> carbon (C=C)	14	I	iodine
3	C	sp <sup>2</sup> carbon (C=O)	15	S	sulfide (-S-)
4	C	sp carbon	16	S+	sulfonium
5	H	hydrogen (see others)	17	S	sulfoxide (use S=O)
6	O	oxygen (single bonded)	18	S	sulfone (use two S=O)
7	O	oxygen (double bonded)	19	Si	silane
8	N	sp <sup>3</sup> nitrogen	20	LP	lone pair of electrons
9	N	sp <sup>2</sup> nitrogen	21	H	hydroxyl hydrogen
10	N	sp nitrogen	22	C	cyclopropane carbon
11	F	fluorine	23	H	amine hydrogen
12	Cl	chlorine	24	H	carboxylic acid hydrogen

# Successes of molecular mechanics

- Calculations are fast
- Geometry optimisations of small and medium-size molecules can be run on a PC
- Conformations of macromolecules can be calculated using workstations
- Reasonable geometries are usually obtained
- Provides input structure for more elaborated calculations

2003-01-31

## In summary

- *Ab initio* and semi-empirical methods can be used to calculate molecular properties that depend on the electron distribution
- Molecular mechanics is a purely empirical approach, but perform well in the field of structure and geometry
- Elaborate computation approaches are not necessarily more accurate, since complex models with many parameters are often more restricted (constrained to a limited domain)
- Different computational approaches are optimal for different research issues

# Demonstrations with HyperChem®

- Structural properties (geometry) of cyclohexane
- Distorting the bond length
- Conformation and rotation barrier of 2,2',4,4',6,6'-hexachlorobiphenyl

2003-01-31