Molecular Mechanics

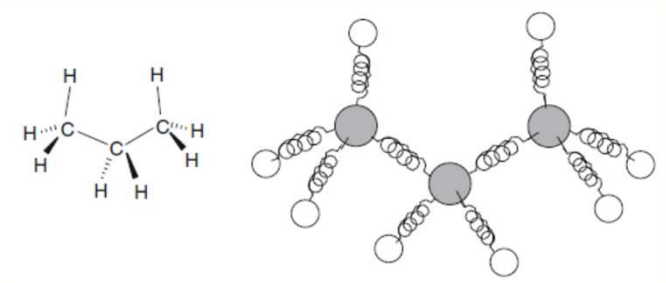
Chapter 3

Abstract

- Molecular mechanics (MM) rests on a view of molecules as balls held together by springs.
- The potential energy of a molecule can be written as the sum of terms involving bond stretching, angle bending, dihedral angles and nonbonded interactions.
- Giving these terms explicit mathematical forms constitutes devising a forcefield, and giving actual numbers to the constants in the forcefield constitutes parameterizing the field.
- An example is given of the devising and parameterization of an MM forcefield.
- Calculations on biomolecules is a very important application of MM, and the pharmaceutical industry designs new drugs with the aid of MM.
- Organic synthesis now makes considerable use of MM, which enables chemists to estimate which products are likely to be favored and to devise more realistic routes to a target molecule.
- In molecular dynamics MM is used to generate the forces acting on molecules and hence to calculate their motions.

Perspective

- Molecular mechanics (MM) is based on a mathematical model of a molecule as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds).
- Within the framework of this model, the energy of the molecule changes with geometry because the springs resist being stretched or bent away from some “natural” length or angle, and the balls resist being pushed too closely together.
- The MM model clearly ignores electrons.



- The principle: express the energy of a molecule as a function of its resistance toward bond stretching, bond bending, and atom crowding.
- Use this energy equation to find the bond lengths, angles, and dihedrals corresponding to the minimum-energy geometries (locals on PES).

- The mathematical expression for the energy, with the parameters in it, constitute a forcefield (are sometimes called forcefield methods).
- Forcefield?: the negative of the first derivative of PE of a particle with respect to displacement along some direction is the force on the particle;
- a “forcefield” E(x, y, z coordinates of atoms) can be differentiated to give the force on each atom.

- The method makes no reference to electrons, and so cannot (except by some kind of empirical algorithm) throw light on electronic properties like charge distributions or nucleophilic and electrophilic behaviour.

- MM implicitly uses the Born–Oppenheimer approximation.
- An important point is that the concept of a bond is central to MM.
- To do a MM calculation you must specify each bond as single, double, etc., since this tells the program how strong a bond to use.

- In an electronic structure calculation–ab initio, semiempirical, and DFT – a molecule is defined by the relative positions of its atomic nuclei, the charge, and the “multiplicity” (which follows easily from the number of unpaired electrons).

- An oxygen nucleus and two protons with the right x, y, z coordinates, enough electrons for no charge, and multiplicity one (no unpaired electrons) is a water molecule.
- There is no need to mention bonds here, although the chemist might wish to somehow extract this useful concept from this picture of nuclei and electrons.
- This can be done by calculating the electron density and associating a bond with, for example, a path along which electron density is concentrated, but there is no unique definition of a bond in electronic structure theory.
- In some graphical interfaces, bonds are specified by the user, while in others they are shown by the program depending on the separation of pairs of atoms.

- Some MM programs: MM1, MM2 and MM3, MM4, Sybyl and UFF .
- MM programs like Sybyl and UFF will handle molecules involving much of the periodic table.

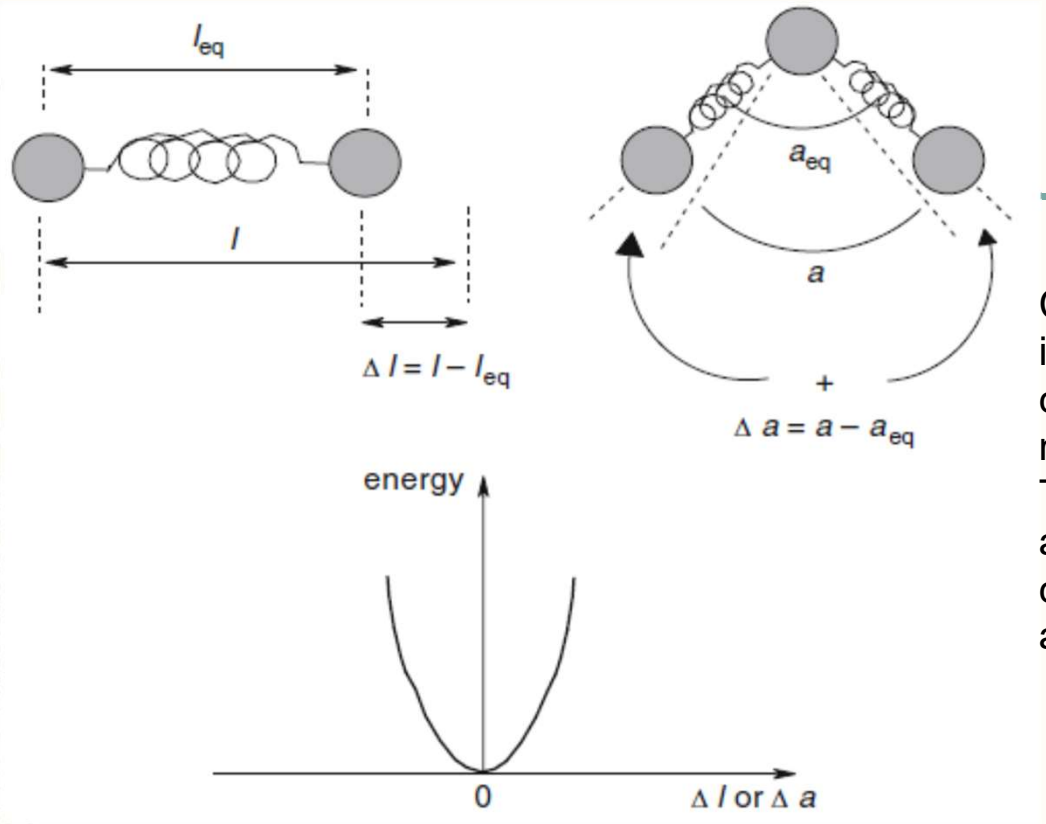
- MM is the most widely-used method for computing the geometries and energies of large biological molecules like proteins and nucleic acids
- Although, recently, semiempirical and even ab initio methods have begun to be applied to these large molecules.

The Basic Principles of Molecular Mechanics 1-Developing a Forcefield

- The potential energy of a molecule can be written

$$E = \sum_{bonds} E_{stretch} + \sum_{angles} E_{bend} + \sum_{dihedrals} E_{torsion} + \sum_{pairs} E_{nonbond}$$

- where $E_{stretch}$ etc. are energy contributions from bond stretching, angle bending, torsional motion (rotation) around single bonds, and interactions between atoms or groups which are nonbonded (not directly bonded together).
- The sums are over all the bonds, all the angles, all the dihedral angles, and all pairs of significant nonbonded interactions.
- The mathematical form of these terms and the parameters in them constitute a particular forcefield.



Changes in bond lengths or in bond angles result in changes in the energy of a molecule.
The energy is approximately a quadratic function of the change in bond length or angle.

- **The Bond Stretching Term:**

– The increase in the energy of a spring when it is stretched is approximately proportional to the square of the extension:

$$\Delta E_{stretch} = k_{stretch}(l - l_{eq})^2$$

– $k_{stretch}$ = the proportionality constant (actually one-half the force constant of the spring or bond.

– l = length of the bond when stretched.

– l_{eq} = equilibrium length of the bond.

– If we take the energy corresponding to the l_{eq} as the zero of energy, we can replace $\Delta E_{stretch}$ by $E_{stretch}$:

$$E_{stretch} = k_{stretch}(l - l_{eq})^2$$

- **The Angle Bending Term:** The increase in energy of system ball-spring-ball-spring- ball is approximately proportional to the square of the increase in the angle:

$$E_{bend} = k_{bend}(a - a_{eq})^2$$

- k_{bend} = a proportionality constant (one-half the angle bending force constant);
- a = size of the angle when distorted.
- a_{eq} = equilibrium size of the angle.

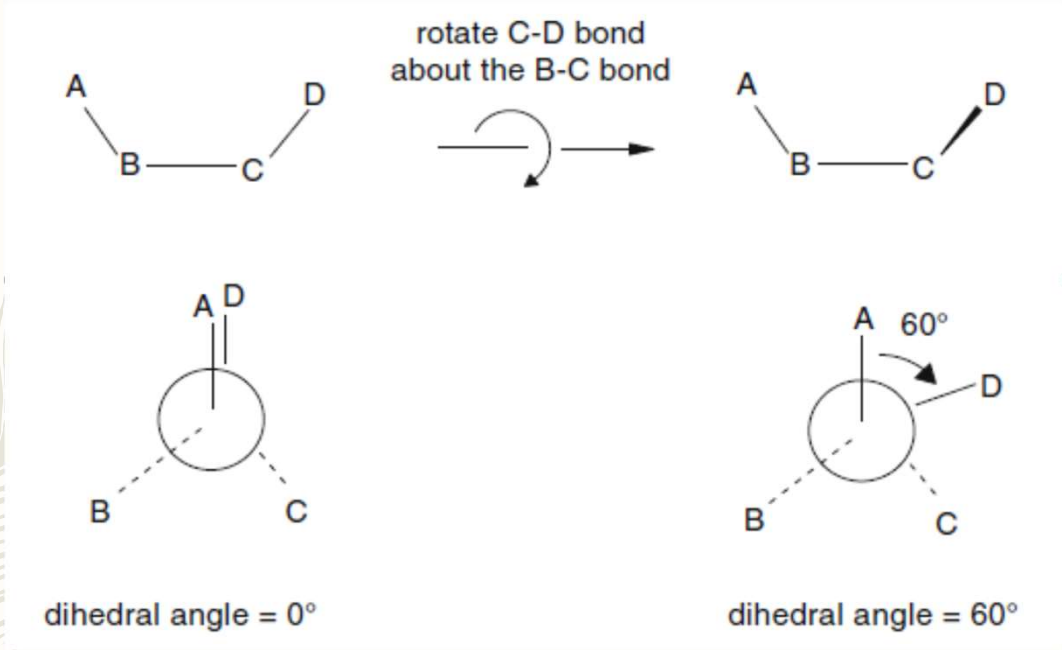
- **The Torsional Term:**

– Consider four atoms sequentially bonded: A–B–C–D. The dihedral angle or torsional angle of the system is the angle between the A–B bond and the C–D bond as viewed along the B–C bond.

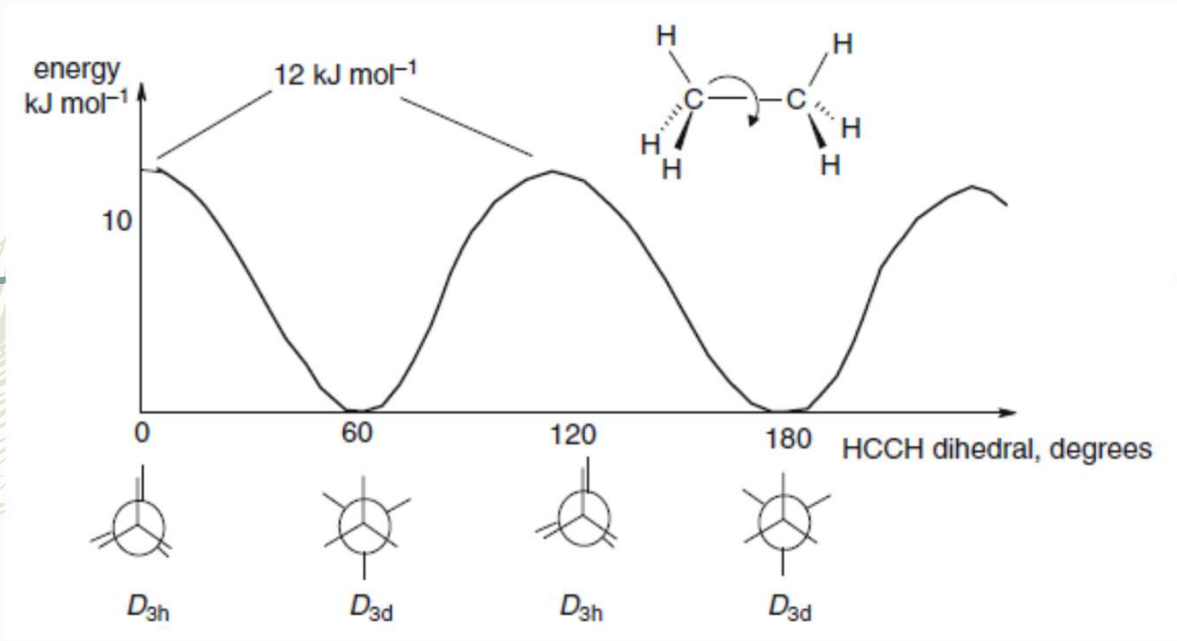
– Conventionally this angle is considered positive if regarded as arising from clockwise rotation of the back bond (C–D) with respect to the front bond (A–B).

– Since the geometry repeats itself every 360, the energy varies with the dihedral angle in a sine or cosine pattern

$$E_{torsion} = k_o + \sum_{r=1}^{n} k_r[1 + \cos(r\theta)]$$



Dihedral angles (torsional angles) affect molecular geometries and energies. The energy is a periodic (cosine or combination of cosine functions) function of the dihedral angle.



Variation of the energy of ethane with dihedral angle. The curve can be represented as a cosine function.

- **The Nonbonded Interactions Term:**

– This represents the change in potential energy with distance apart of atoms A and B that are not directly bonded (as in A–B) and are not bonded to a common atom (as in A–X–B).

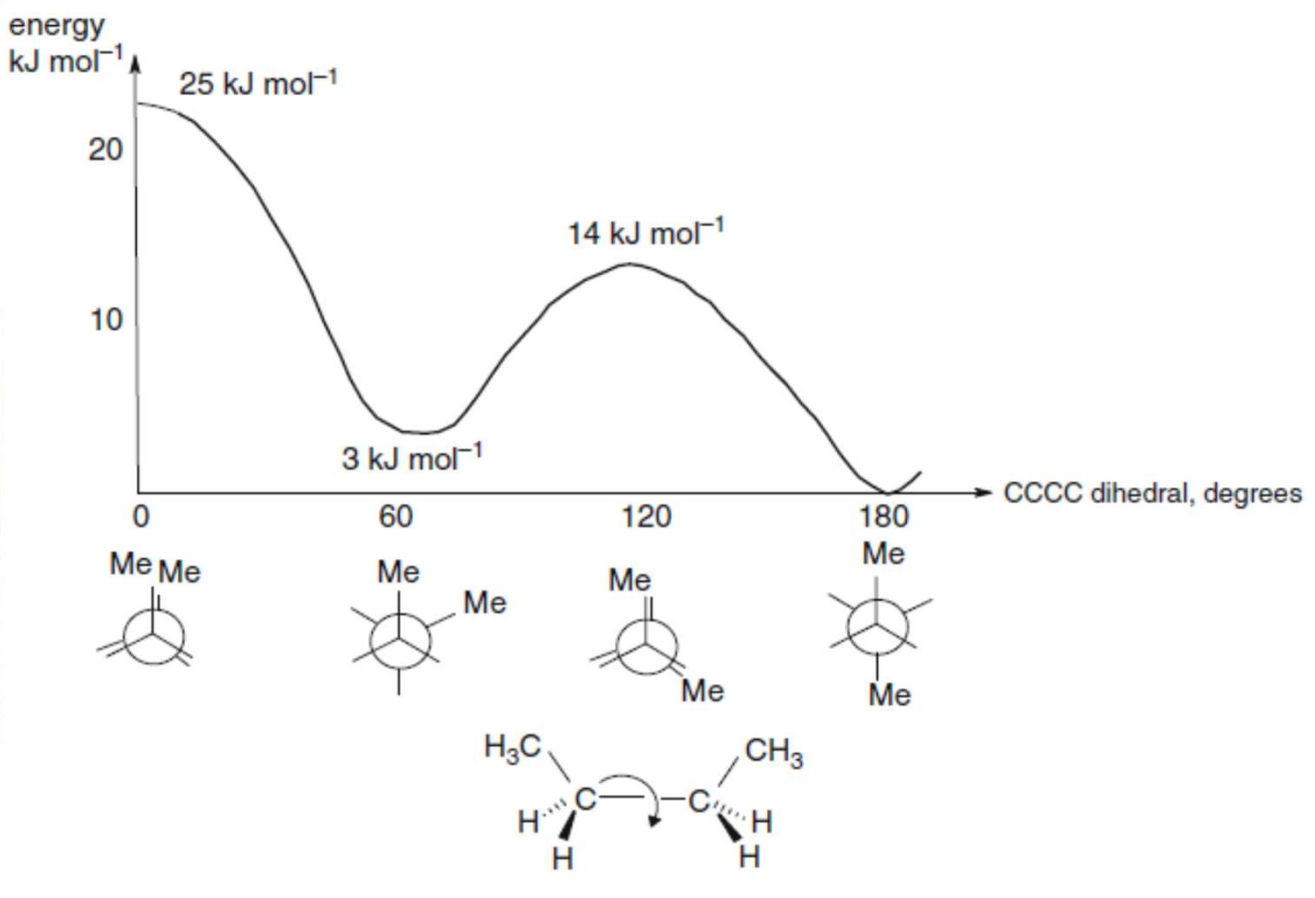
– These atoms, separated by at least two atoms (A–X–Y–B) or even in different molecules.

– Note: that the A-B case is accounted for by the term $E_{stretch}$, and the A–X–B term by E_{bend},

– The nonbonded term $E_{nonbond}$ is, for the A–X–Y–B case, superimposed upon the torsional term $E_{torsion}$.

$$E_{nonbond} = k_{nb}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

so-called Lennard-Jones 12–6 potentia



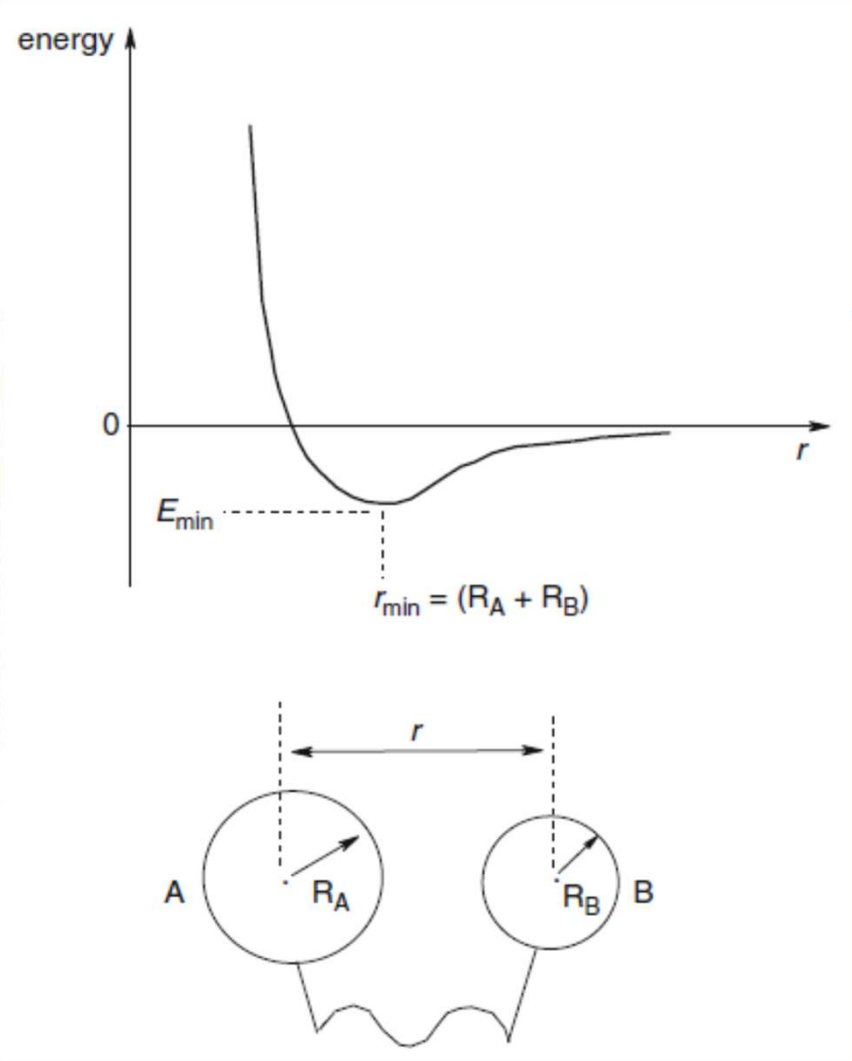
Variation of the energy of butane with dihedral angle. The curve can be represented by a sum of cosine functions

- r = the distance between the centers of the nonbonded atoms or groups.
- The function reproduces the small attractive dip in the curve (represented by the negative term) as the atoms or groups approach one another, then the very steep rise
- in potential energy (represented by the positive, repulsive term raised to a largepower) as they are pushed together closer than their van der Waals radii. Setting dE/dr = 0, we find that for the energy minimum in the curve the corresponding value of r is r_{min} = 21=6s,

$$\text{i.e.} \quad \sigma = 2^{-1/6} r_{\min}$$

- If we assume that this minimum corresponds to van der Waals contact of the nonbonded groups, then $r_{min} = (R_A + R_B)$, the sum of the van der Waals radii of the groups A and B. So

$$2^{1/6}\sigma = (R_A + R_B)$$



Variation of the energy of a molecule with separation of nonbonded atoms or groups. Atoms/groups A and B may be in the same molecule (as indicated here) or the interaction may be intermolecular. The minimum energy occurs at van der Waals contact. For small nonpolar atoms or groups the minimum energy point represents a drop of a few kJ mol^{-1} $_1$ (E_{min}=1.2 kJ mol^{-1} $_1$ for CH_4/CH_4), but short distances can make nonbonded interactions destabilize a molecule by many kJ mol^{-1}

- and so

$$\sigma = 2^{-1/6}(R_A + R_B) = 0.89\ (R_A + R_B)$$

- Thus σ can be calculated from r_{min} or estimated from the van der Waals radii.
- Setting E = 0, we find that for this point on the curve r = σ,

$$\sigma = r(E = 0)$$

- If we set r = r_{min} [illegible] Eq. 3.5, we find

$$E(r = r_{min}) = (-1/4)k_{nb} \qquad k_{nb} = -4E(r = r_{min})$$

- So k_{nb} can be calculated from the depth of the energy minimum.
- In deciding to use equations of the form (3.2), (3.3), (3.4) (3.5) we have decided on a particular MM forcefield. There are many alternative forcefields. For example, we might have chosen to approximate $E_{stretch}$ by the sum of a quadratic and a cubic term:

$$E_{stretch} = k_{stretch}(l - l_{eq})^2 + k(l - l_{eq})^3$$

- This gives a somewhat more accurate representation of the variation of energy with length. Again, we might have represented the nonbonded interaction energy by a more complicated expression than the simple 12–6 potential of Eq. 3.5 (which is by no means the best form for nonbonded repulsions).
- Such changes would represent changes in the forcefield.

Parameterizing a Forcefield

- We can now consider putting actual numbers, $k_{stretch}$, etc., into following Eqs to give expressions that we can actually use.

$$E_{stretch} = k_{stretch}(l - l_{eq})^2$$

$$E_{bend} = k_{bend}(a - a_{eq})^2$$

$$E_{torsion} = k_o + \sum_{r=1}^{n} k_r[1 + \cos(r\theta)]$$

$$E_{nonbond} = k_{nb}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

- The process of finding these numbers is called parameterization (or parametrizing) the forcefield. The set of molecules used for parameterization, perhaps 100 for a good forcefield, is called the training set.
- In the purely illustrative example below we use just ethane, methane and butane.

- **Parameterizing the Bond Stretching Term**
- A forcefield can be parameterized by reference to experiment (empirical parameterization) or by getting the numbers from high-level ab initio or density functional calculations, or by a combination of both approaches.

- For the bond stretching term, we need $k_{stretch}$ and l_{eq}.
- Experimentally:

- $k_{stretch}$: from IR spectra, (the stretching frequency of a bond depends on the force constant and the masses of the atoms involved),
- l_{eq} : from X-ray diffraction, electron diffraction, or microwave spectroscopy.

$k_{stretch}$ for the C/C bond of ethane by ab initio calculations:

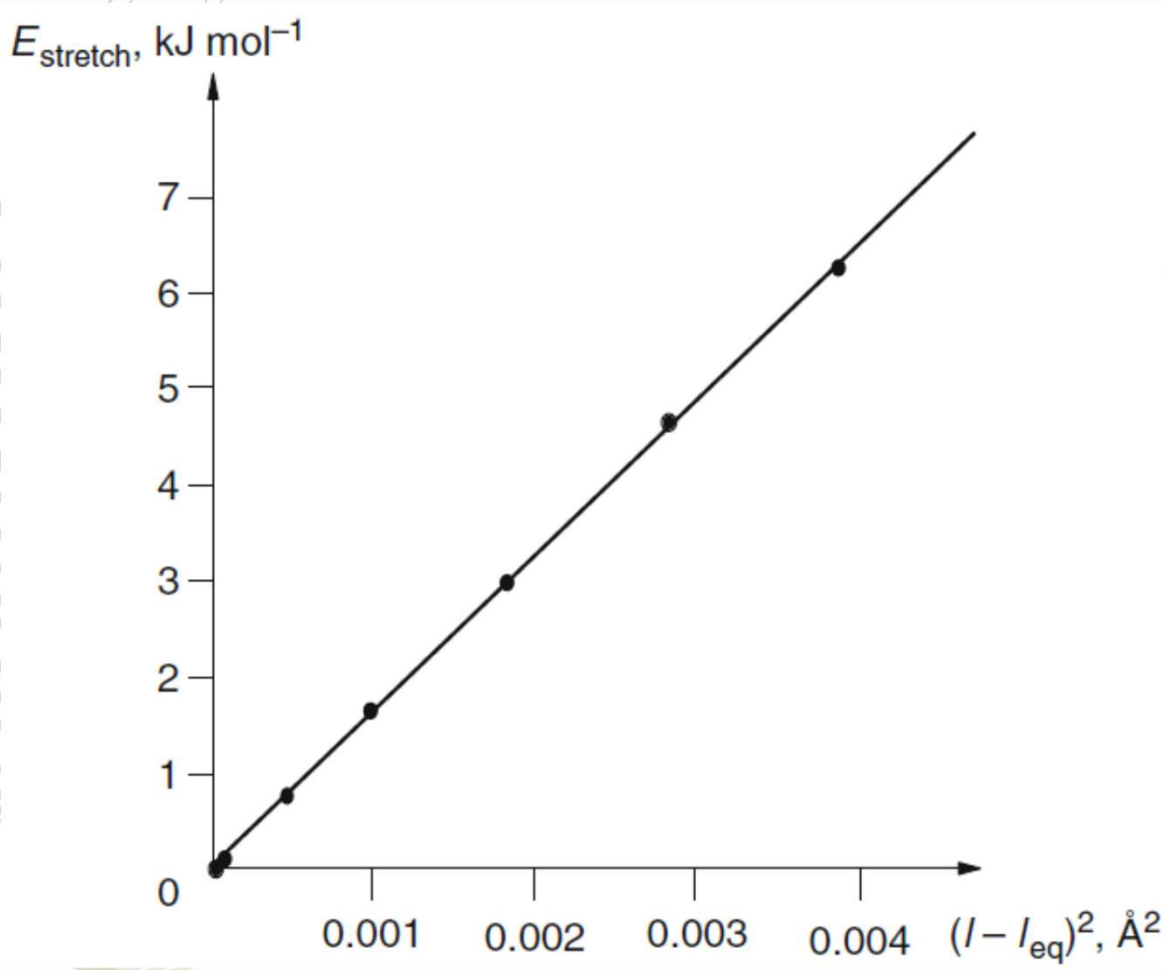
- Normally high-level ab initio calculations would be used but for illustrative purposes we can use the low-level but fast STO-3G method.

$$E_{stretch} = k_{stretch}(l - l_{eq})^2$$

- a plot of $E_{stretch}$ against $(l–l_{eq})^2$ should be linear with a slope of $k_{stretch}$.

Change in energy as the C–C bond in CH_3–CH_3 is stretched away from its equilibrium length. The calculations are ab initio (STO-3G). Bond lengths are in A

C–C length, l	$l - l_{eq}$	$(l - l_{eq})^2$	$E_{stretch}$, kJ mol^{-1}
1.538	0	0	0
1.550	0.012	0.00014	0.29
1.560	0.022	0.00048	0.89
1.570	0.032	0.00102	1.86
1.580	0.042	0.00176	3.15
1.590	0.052	0.00270	4.75
1.600	0.062	0.00384	6.67



Energy vs. the square of the extension of the C–C bond in CH_3–CH_3.

$$l_{eq}(\mathrm{C-C}) = 1.538\ \text{Å}$$

$$k_{stretch}(\mathrm{C-C}) = 1{,}735\ \mathrm{kJ\ mol^{-1}\ Å^{-2}}$$

– Similarly, the CH bond of methane

$$E_{stretch} = k_{stretch}(l - l_{eq})^2$$

$$\begin{cases} l_{eq'}(\mathrm{C-H}) = 1.083\text{Å} \\ k_{stretch}(\mathrm{C-H})1.934\ \mathrm{kJ\ mol^{-1}\ Å^{-2}} \end{cases}$$

– Parameterizing the Angle Bending Term

$$E_{bend} = k_{bend}(a - a_{eq})^2 \longrightarrow$$

$$\begin{cases} a_{eq}(\mathrm{HCC}) = 110.7^\circ \\ k_{bend}(\mathrm{HCC}) = 0.093\,\mathrm{kJ\,mol^{-1}deg^{-2}} \end{cases}$$

$$\begin{cases} a_{eq}(\mathrm{CCC}) = 112.5^\circ \\ k_{bend}(\mathrm{CCC}) = 0.110\,\mathrm{KJ\,mol^{-1}deg^{-2}} \end{cases}$$

Parameterizing the Torsional Term

- For the ethane case, the equation for energy as a function of dihedral angle can be deduced fairly simply by adjusting the basic equation E = cos θ to give

$$E = 1/2E_{max}[1 + \cos 3(\theta + 60)]$$

- For butane: experimenting with a curve-fitting program shows that a reasonably accurate torsional potential energy function can be created with five parameters, k_0 and k_1–k_4:

$$E_{torsion}(CH_3CH_2 - CH_2CH_3) = k_0 + \sum_{r=1}^{4} k_r[1 + \cos(r\theta)]$$

– The experimental potential energy values for rotation about the central C–C bond of CH3CH2–CH2CH3 can be approximated by $E_{torsion}$(CH3CH2-CH2CH3)= $k_0 + \sum_{r=1}^{4} k_r[1 + \cos(r\theta)]$ with k_0 = 20.1, k_1 = 4.7, k_2 = 1.91, k_3 = 7.75, k_4 = 0.58. Experimental energy values at 30, 90 , and 150 were interpolated from those at 0, 60 , 120, and 180; energies are in kJ mol^{-1}

θ (deg)	*E* (calculated)	*E* (experimental)
0	0.15	0
30	6.7	7.0
60	14	14
90	8.8	9.0
120	3.5	3.3
150	15	15
180	25	25

Parameterizing the Nonbonded Interactions Term

- To parameterize

$$E_{nonbond} = k_{nb}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

- we might perform ab initio calculations in which the separation of two atoms or groups in different molecules (to avoid the complication of concomitant changes in bond lengths and angles) is varied, and fit Eq. 3.5 to the energy vs. distance results. For nonpolar groups this would require quite high-level calculations, as van der Waals or dispersion forces are involved. We shall approximate the nonbonded interactions of methyl groups by the interactions of methane molecules, using experimental values of k_{nb} and σ, derived from studies of the viscosity or the compressibility of methane. The two methods give slightly different values, but we can use the values

$$k_{nb} = 4.7\,\text{kJ mol}^{-1} \qquad \sigma = 3.85\ \text{Å}$$

Summary of the Parameterization of the Forcefield Terms

$$E_{stretch}(\mathrm{C-C}) = 1735(l - 1.538)^2$$

$$E_{stretch}(\mathrm{C-H}) = 1934(l - 1.083)^2$$

$$E_{bend}(\mathrm{HCH}) = 0.093(a - 110.7)^2$$

$$E_{bend}(\mathrm{CCC}) = 0.110(a - 112.5)^2$$

$$E_{torsion}(\mathrm{CH_3CCCH_3}) = k_0 + \sum_{r=1}^{4} k_r[1 + \cos(r\theta)]$$

$$E_{nonbond}(CH_3/CH_3) = 4.7\left[\left(\frac{3.85}{r}\right)^{12} - \left(\frac{3.85}{r}\right)^{6}\right]$$

- This parameterization is only illustrative of the principles involved;
- An accurate, practical forcefield would be parameterized as a best fit to many experimental and/or calculational results, and would have different parameters for different kinds of bonds, e.g. C–C for acyclic alkanes, for cyclobutane and for cyclopropane.

- A forcefield able to handle not only hydrocarbons would obviously need parameters involving elements other than hydrogen and carbon.
- Practical forcefields also have different parameters for various atom types, like sp^3 carbon vs. sp^2 carbon, or amine nitrogen vs. amide nitrogen. a different value would be used for, say, stretching involving an sp^3/sp^3 C–C bond than for an sp^2/sp^2 C–C bond.
- This is clearly necessary since the force constant of a bond depends on the hybridization of the atoms involved
- Some forcefields account for the variation of bond order with conformation by performing a simple PPP MO calculation to obtain the bond order.

- A sophisticated forcefield might also consider H/H nonbonded interactions explicitly, rather than simply subsuming them into methyl/methyl interactions
- nonbonding interactions between polar groups need to be accounted for in a field not limited to hydrocarbons. usually by the well-known potential energy/electrostatic charge relationship

$$E = k(q_1 q_2 / r)$$

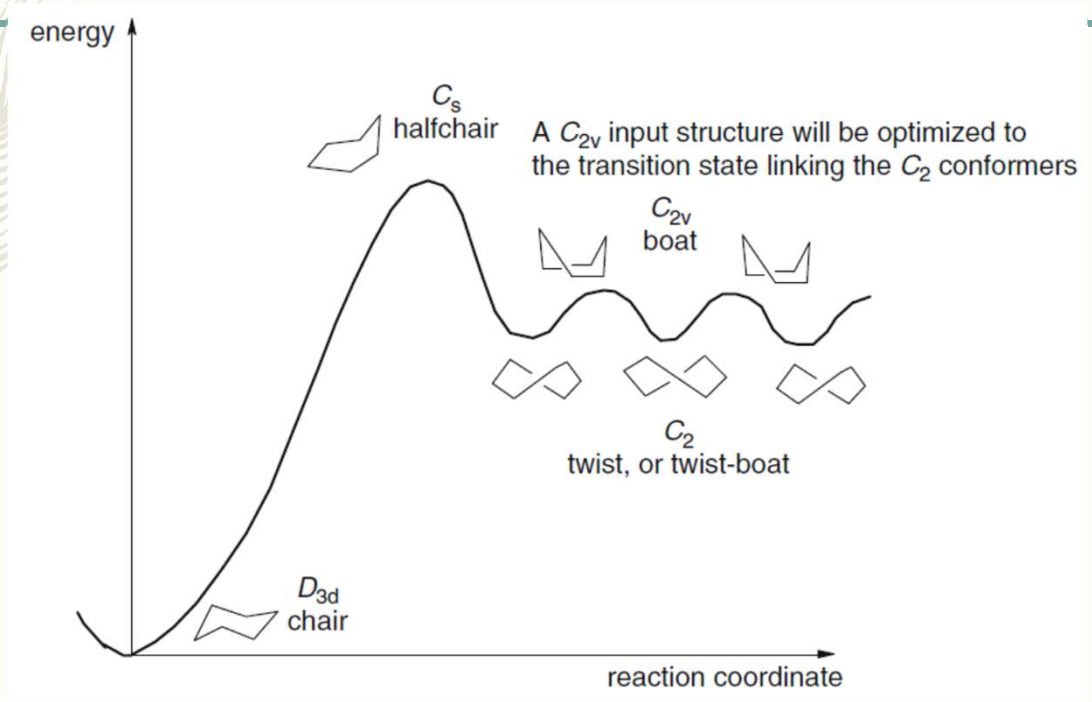
- which has also been used to model hydrogen bonding

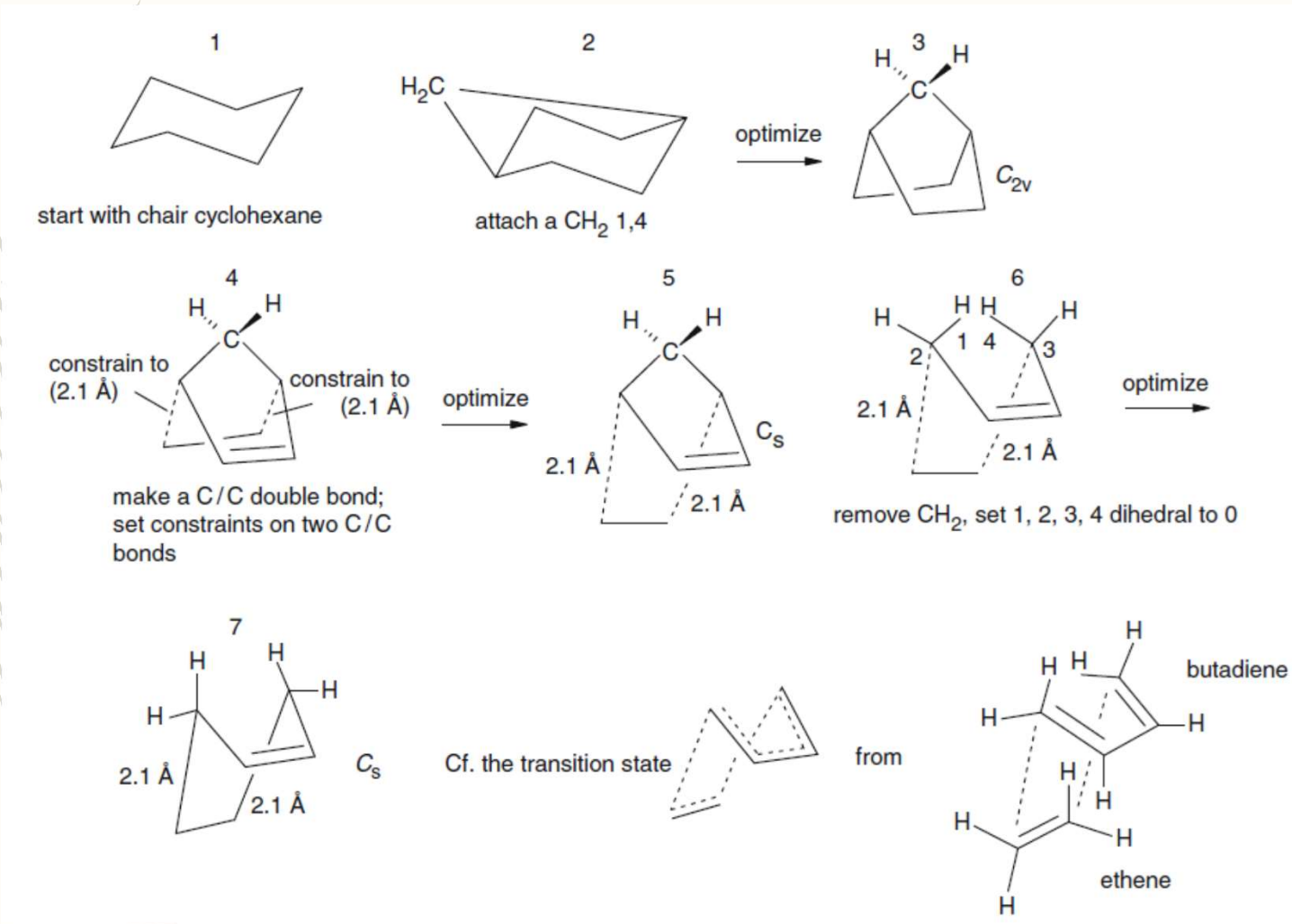
- A subtler problem: stretching, bending, torsional and nonbonded terms are not completely independent.
- For example, the butane torsional potential energy curve does not apply precisely to all CH3–C–C–CH3 systems (barrier heights will vary with the length of the central C–C bond: decrease in the bond length → decrease in the interactions between the CH3's and H's on two carbons of central C–C.
- So, the k's in the $E_{torsion}$ are also a function of X–Y length.
- Actually, partitioning the energy of a molecule into stretching, bending, etc. terms is somewhat formal (torsional barrier in butane can be considered to be partly due to nonbonded interactions between the methyl groups)
- It should be realized that there is no one, right functional form for an MM forcefield; accuracy, versatility and speed of computation are the deciding factors in devising a forcefield.

Examples of the Use of Molecular Mechanics

- the main applications of MM:
- 1. To obtain reasonable input geometries for lengthier kinds of calculations.
- 2. To obtain good geometries (and perhaps energies) for small- to medium-sized molecules.
- 3. To calculate the geometries and energies of very large molecules, usually polymeric biomolecules (proteins and nucleic acids).
- 4. To generate the potential energy function under which molecules move, for molecular dynamics or Monte Carlo calculations.
- 5. As a (usually quick) guide to the feasibility of, or likely outcome of, reactions in organic synthesis.

To Obtain Reasonable Input Geometries for Lengthier (Ab Initio, Semiempirical or Density Functional) Kinds of Calculations





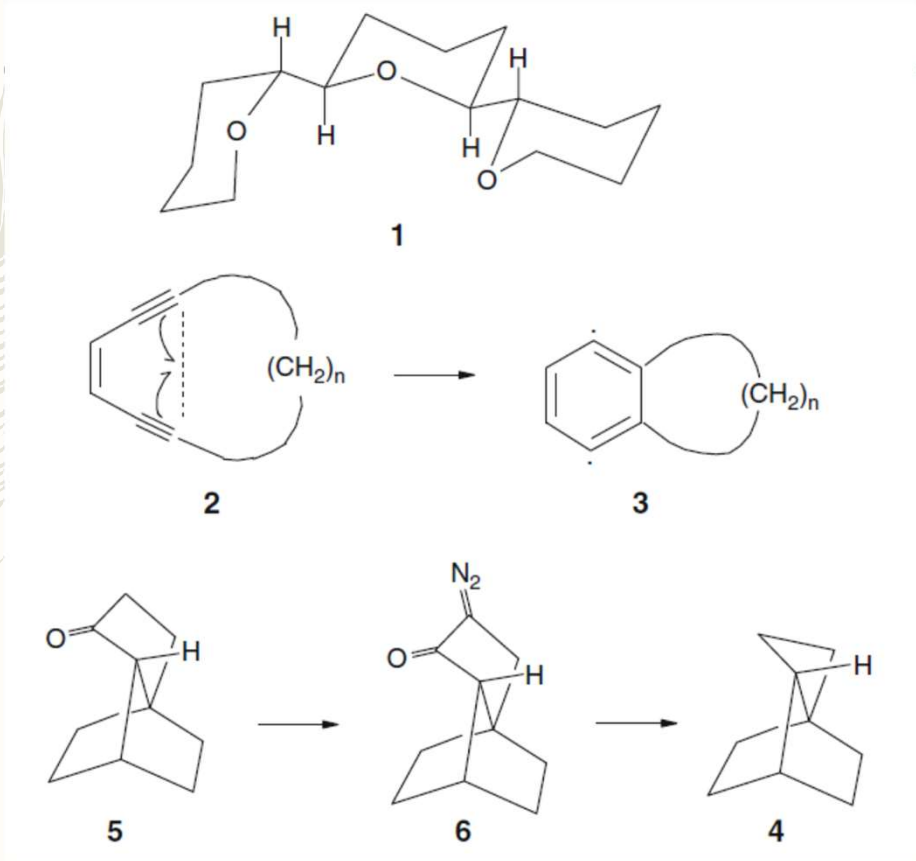
Using molecular mechanics to get the (approximate) transition state for the Diels–Alder reaction of butadiene with ethene. This procedure gives a structure with the desirable C_s, rather than a lower, symmetry

To Obtain Good Geometries (and Perhaps Energies) for Small- to Medium-Sized Molecules

To Calculate the Geometries and Energies of Very Large Molecules, Usually Polymeric Biomolecules (Proteins and Nucleic Acids)

To Generate the Potential Energy Function Under Which Molecules Move, for Molecular Dynamics or Monte Carlo Calculations

As a (Usually Quick) Guide to the Feasibility of, or Likely Outcome of, Reactions in Organic Synthesis



synthesiz with the aid of molecular mechanics

Frequencies and Vibrational Spectra Calculated by MM

Strengths and Weaknesses of Molecular Mechanics

- **Strengths**
- MM is fast. is not always at the expense of accuracy (for the kinds of molecules for which it has been parameterized). MM is undemanding in its hardware requirements (on standard personal computers are quite practical).
- speed, (frequent) accuracy and modest computer requirements → have given MM a place in many modelling programs.
- speed and the availability of parameters for almost all the elements → can supply reasonably good input geometries for semiempirical, ab initio or density functional calculations (even when it does not provide very accurate geometries)
- recent ability of MM programs to calculate IR spectra with some accuracy
- Unfortunately, MM can't be used as a shortcut to obtaining frequencies for a species optimized by a quantum mechanical calculation, since frequencies must be calculated using the same method used for the geometry optimization.

- **Weaknesses**
- The weaknesses stem from the fact that it ignores electrons.
- The philosophy behind MM → a molecule is a collection of atoms subject to forces and using forces to express the energy in terms of the geometric parameters.
- By parameterization MM can “calculate” electronic properties (using bond dipoles it can find a dipole moment for a molecule, and using values that have been calculated for various atom types by quantum mechanics it can assign charges to atoms. However, such results are obtained purely by analogy, and their reliability can be negated by unexpected electronic factors to which MM is oblivious.
- MM cannot provide information about the shapes and energies of molecular orbitals nor about related phenomena such as electronic spectra.

- Because of the severely empirical nature of MM, interpreting MM parameters in terms of traditional physical concepts is dangerous; for example, the bond-stretching and angle-bending parameters cannot rigorously be identified with spectroscopic force constants
- Other dangers in using MM are:

➢ 1. Using an inappropriate forcefield: a field parameterized for one class of compounds is not likely to perform well for other classes.

➢ 2. Transferring parameters from one forcefield to another. This is usually not valid.

➢ 3. Optimizing to a stationary point that may not really be a minimum (it could be a “maximum”, a transition state).

- 4. MM programs, more so than semiempirical ones and unlike ab initio or DFT programs, are ruled by empirical factors
- 5. Ignoring solvent and nearby ions: for polar molecules using the in vacuo structure can lead to quite wrong geometries and energies.
- 6. Lack of caution about comparing energies calculated with MM. The method calculates the energy of a molecule relative to a hypothetical strainless idealization of the molecule.

- Using MM to calculate the relative energy of two isomers by comparing their strain energies (the normal MM energies) is dangerous because the two strain energies are not necessarily relative to the same hypothetical unstrained species (strain energies are not an unambiguous observable). This is particularly true for functional group isomers, like (CH3)2O/CH3CH2OH and CH3COCH3/H2C=C(OH)CH3, which have quite different atom types. For isomers consisting of the same kinds of atoms (alkanes cf. alkanes, say), and especially for conformational isomers and E/Z isomers (geometric isomers), a good MM forcefield should give strain energies which reasonably represent relative enthalpies. For example, the MMFF gives for $CH_3COCH_3/H_2C{=}C(OH)CH_3$ strain energies of 6.9/–6.6 kJ mol^{-1}, i.e. relative energies of 0/–13 kJ mol^{-1}, but the experimental value is ca. 0/44 kJ mol–1, i.e. H2C=C(OH)CH3 is much the higher-energy molecule. On the other hand, the MMFF yields for gauche–butane/anti-butane strain energies of –21.3/–18.0 kJ mol^{-1}, i.e. relative energies of 0/3.3 kJ mol^{-1}, reasonably close to the experimental value of 0/2.8 kJ mol–1. For chair (D2d), twist (D2), and boat (C2v) cyclohexane, the MMFF strain energies are –14.9, 9.9 and 13.0 kJ mol^{-1}, i.e. relative energies of 0, 24.8 and 27.9 kJ mol^{-1}, cf. the experimental the estimates of 0, 24 and 29 kJ mol^{-1}.
- MM programs can be parameterized to give, not just strain energy, but enthalpies of formation. Although chemists often compare stabilities of isomers using enthalpies, we should remember that equilibria are actually determined by free energies.

- 7. Assuming that the major conformation determines the product. In fact, in a mobile equilibrium the product ratio depends on the relative reactivities, not relative amounts, of the conformers (the Curtin-Hammett principle).
- 8. Failure to exercise judgement: small energy differences (say up to 10–20 kJ Mol^{-1}) mean nothing in many cases.

Summary

- This chapter explains the basic principles of molecular mechanics (MM), which rests on a view of molecules as balls held together by springs.
- The potential energy of a molecule can be written as the sum of terms involving bond stretching, angle bending, dihedral angles and nonbonded interactions.
- Giving these terms explicit mathematical forms constitutes devising a forcefield, and giving actual numbers to the constants in the forcefield constitutes parameterizing the field.

- MM is widely used to create reasonable geometries for input to other calculations.
- Such calculations are fast and can be very accurate, provided that the forcefield has been carefully parameterized for the types of molecules under study.

- Calculations on biomolecules is a very important application of MM (docking and the related aspect of QSAR)
- MM is of some limited use in calculating the geometries and energies of transition states.
- Organic synthesis now makes considerable use of MM, which enables chemists to estimate which products are likely to be favored
- In molecular dynamics MM is used to generate the forces acting on molecules and hence to calculate their motions, and in Monte Carlo simulations MM is used to calculate the energies of the many randomly generated states.

- MM is fast, it can be accurate, it is undemanding of computer power, and it provides reasonable starting geometries for quantum mechanical calculations.
- It ignores electrons, and so can provide parameters like dipole moment only by analogy.

- One must be cautious about the applicability of MM parameters to the problem at hand.
- Stationary points from MM, even when they are relative minima, may not be global minima.
- Ignoring solvent effects can give erroneous results for polar molecules.
- MM gives strain energies, the difference of which for structurally similar isomers represent enthalpy differences; parameterization to give enthalpies of formation is possible.
- Strictly speaking, relative amounts of isomers depend on free energy differences.
- The major conformation (even when correctly identified) is not necessarily the reactive one.