

Empirical Force Field Models: Molecular Mechanics

4.1 Introduction

Many of the problems that we would like to tackle in molecular modelling are unfortunately too large to be considered by quantum mechanics. Quantum mechanical methods deal with the electrons in a system, so that even if some of the electrons are ignored (as in the semi-empirical schemes) a large number of particles must still be considered, and the calculations are time-consuming. Force field methods (also known as molecular mechanics) ignore the electronic motions and calculate the energy of a system as a function of the nuclear positions only. Molecular mechanics is thus invariably used to perform calculations on systems containing significant numbers of atoms. In some cases force fields can provide answers that are as accurate as even the highest-level quantum mechanical calculations, in a fraction of the computer time. However, molecular mechanics cannot of course provide properties that depend upon the electronic distribution in a molecule.

That molecular mechanics works at all is due to the validity of several assumptions. The first of these is the Born–Oppenheimer approximation, without which it would be impossible to contemplate writing the energy as a function of the nuclear coordinates at all. Molecular mechanics is based upon a rather simple model of the interactions within a system with contributions from processes such as the stretching of bonds, the opening and closing of angles and the rotations about single bonds. Even when simple functions (e.g. Hooke's law) are used to describe these contributions the force field can perform quite acceptably. Transferability is a key attribute of a force field, for it enables a set of parameters developed and tested on a relatively small number of cases to be applied to a much wider range of problems. Moreover, parameters developed from data on small molecules can be used to study much larger molecules such as polymers.

4.1.1 A Simple Molecular Mechanics Force Field

Many of the molecular modelling force fields in use today for molecular systems can be interpreted in terms of a relatively simple four-component picture of the intra- and inter-molecular forces within the system. Energetic penalties are associated with the deviation of bonds and angles away from their 'reference' or 'equilibrium' values, there is a function

that describes how the energy changes as bonds are rotated, and finally the force field contains terms that describe the interaction between non-bonded parts of the system. More sophisticated force fields may have additional terms, but they invariably contain these four components. An attractive feature of this representation is that the various terms can be ascribed to changes in specific internal coordinates such as bond lengths, angles, the rotation of bonds or movements of atoms relative to each other. This makes it easier to understand how changes in the force field parameters affect its performance, and also helps in the parametrisation process. One functional form for such a force field that can be used to model single molecules or assemblies of atoms and/or molecules is:

$$\begin{aligned} \mathcal{V}(\mathbf{r}^N) = & \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\ & + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \end{aligned} \quad (4.1)$$

$\mathcal{V}(\mathbf{r}^N)$ denotes the potential energy, which is a function of the positions (\mathbf{r}) of N particles (usually atoms). The various contributions are schematically represented in Figure 4.1. The first term in Equation (4.1) models the interaction between pairs of bonded atoms, modelled here by a harmonic potential that gives the increase in energy as the bond length l_i deviates from the reference value $l_{i,0}$. The second term is a summation over all valence angles in the molecule, again modelled using a harmonic potential (a valence angle is the angle formed between three atoms A–B–C in which A and C are both bonded to B). The third term in Equation (4.1) is a torsional potential that models how the energy changes as a bond rotates. The fourth contribution is the non-bonded term. This is calculated between all pairs of atoms (i and j) that are in different molecules or that are in

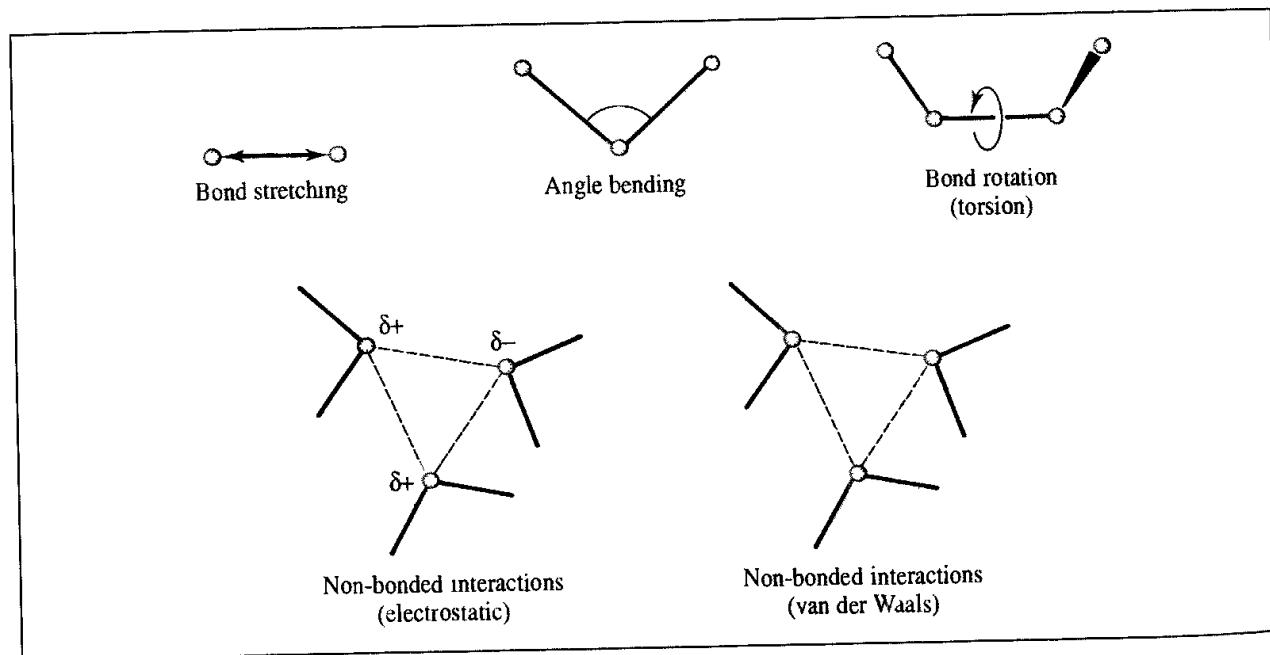


Fig 4.1 Schematic representation of the four key contributions to a molecular mechanics force field bond stretching, angle bending and torsional terms and non-bonded interactions

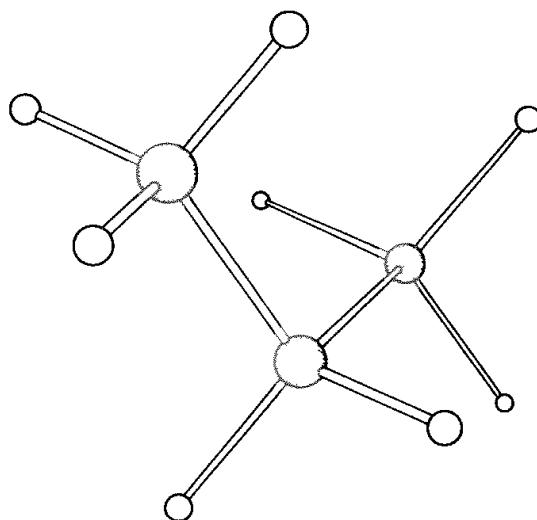


Fig. 4.2. A typical force field model for propane contains ten bond-stretching terms, eighteen angle-bending terms, eighteen torsional terms and 27 non-bonded interactions

the same molecule but separated by at least three bonds (i.e. have a $1, n$ relationship where $n \geq 4$). In a simple force field the non-bonded term is usually modelled using a Coulomb potential term for electrostatic interactions and a Lennard-Jones potential for van der Waals interactions.

We shall discuss the nature of these different contributions in more detail in Sections 4.3–4.10, but here we consider how the simple force field of Equation (4.1) would be used to calculate the energy of a conformation of propane (Figure 4.2). Propane has ten bonds: two C–C bonds and eight C–H bonds. The C–C bonds are symmetrically equivalent but the C–H bonds fall into two classes, one group corresponding to the two hydrogens bonded to the central methylene (CH_2) carbon and one group corresponding to the six hydrogens bonded to the methyl carbons. In some sophisticated force fields different parameters would be used for these two different types of C–H bond, but in most force fields the same bonding parameters (i.e. k_i and $l_{i,0}$) would be used for each of the eight C–H bonds. This is an example of the way in which the same parameters can be used for a wide variety of molecules. There are 18 different valence angles in propane, comprising one C–C–C angle, ten C–C–H angles and seven H–C–H angles. Note that all angles are included in the force field model even though some of them may not be independent of the others. There are 18 torsional terms: twelve H–C–C–H torsions and six H–C–C–C torsions. Each of these is modelled with a cosine series expansion that has minima at the *trans* and *gauche* conformations. Finally, there are 27 non-bonded terms to calculate, comprising 21 H–H interactions and six H–C interactions. The electrostatic contribution would be calculated using Coulomb's law from partial atomic charges associated with each atom and the van der Waals contribution as a Lennard-Jones potential with appropriate ϵ_{ij} and σ_{ij} parameters. A sizeable number of terms are thus included in the force field model, even for a molecule as simple as propane. Even so, the number of terms (73) is many fewer than the number of integrals that would be involved in an equivalent *ab initio* quantum mechanical calculation.

4.2 Some General Features of Molecular Mechanics Force Fields

To define a force field one must specify not only the functional form but also the parameters (i.e. the various constants such as k_i , V_n and σ_{ij} in Equation (4.1)); two force fields may use an identical functional form yet have very different parameters. Moreover, force fields with the same functional form but different parameters, and force fields with different functional forms, may give results of comparable accuracy. A force field should be considered as a single entity; it is not strictly correct to divide the energy into its individual components, let alone to take some of the parameters from one force field and mix them with parameters from another force field. Nevertheless, some of the terms in a force field are sufficiently independent of the others (particularly the bond and angle terms) to make this an acceptable approximation in certain cases.

The force fields used in molecular modelling are primarily designed to reproduce structural properties but they can also be used to predict other properties, such as molecular spectra. However, molecular mechanics force fields can rarely predict spectra with great accuracy (although the more recent molecular mechanics force fields are much better in this regard). A force field is generally designed to predict certain properties and will be parametrised accordingly. While it is useful to try to predict other quantities which have not been included in the parametrisation process it is not necessarily a failing if a force field is unable to do so.

Transferability of the functional form and parameters is an important feature of a force field. Transferability means that the same set of parameters can be used to model a series of related molecules, rather than having to define a new set of parameters for each individual molecule. For example, we would expect to be able to use the same set of parameters for all *n*-alkanes. Transferability is clearly important if we want to use the force field to make predictions. Only for some small systems, where particularly accurate work is required, may it be desirable to develop a model specific to that molecule.

One important point that we should bear in mind as we undertake a deeper analysis of molecular mechanics is that force fields are *empirical*; there is no 'correct' form for a force field. Of course, if one functional form is shown to perform better than another it is likely that form will be favoured. Most of the force fields in common use do have a very similar form, and it is tempting to assume that this must therefore be the optimal functional form. Certainly such models tend to conform to a useful picture of the interactions present in a system, but it should always be borne in mind that there may be better forms, particularly when developing a force field for new classes of molecule. The functional forms employed in molecular mechanics force fields are often a compromise between accuracy and computational efficiency; the most accurate functional form may often be unsatisfactory for efficient computation. As the performance of computers increases so it becomes possible to incorporate more sophisticated models. An additional consideration is that in order to use techniques such as energy minimisation and molecular dynamics, it is usually desirable to be able to calculate the first and second derivatives of the energy with respect to the atomic coordinates.

A concept that is common to most force fields is that of an *atom type*. When preparing the input for a quantum mechanics calculation it is usually necessary to specify the atomic numbers of the nuclei present, together with the geometry of the system and the overall charge and spin multiplicity. For a force field the overall charge and spin multiplicity are not explicitly required, but it is usually necessary to assign an atom type to each atom in the system. The atom type is more than just the atomic number of an atom; it usually contains information about its hybridisation state and sometimes the local environment. For example, it is necessary in most force fields to distinguish between sp^3 -hybridised carbon atoms (which adopt a tetrahedral geometry), sp^2 -hybridised carbons (which are trigonal) and sp -hybridised carbons (which are linear). Each force field parameter is expressed in terms of these atom types, so that the reference angle θ_0 for a tetrahedral carbon atom would be near 109.5° and that for a trigonal carbon would be near 120° . The atom types in some force fields reflect the neighbouring environment as well as the hybridisation and can be quite extensive for some atoms. For example, the MM2, MM3 and MM4 force fields of Allinger and co-workers that are widely used for calculations on 'small' molecules [Allinger 1977; Allinger *et al.* 1989, 1990a, b, 1996a, b; Lii and Allinger 1989; Nevins *et al.* 1996a, b, c] distinguish the following types of carbon atom: sp^3 , sp^2 , sp , carbonyl, cyclopropane, radical, cyclopropene and carbonium ion. In the AMBER force field of Kollman and co-workers [Weiner *et al.* 1984; Cornell *et al.* 1995] the carbon atom at the junction between a six- and a five-membered ring (e.g. in the amino acid tryptophan) is assigned

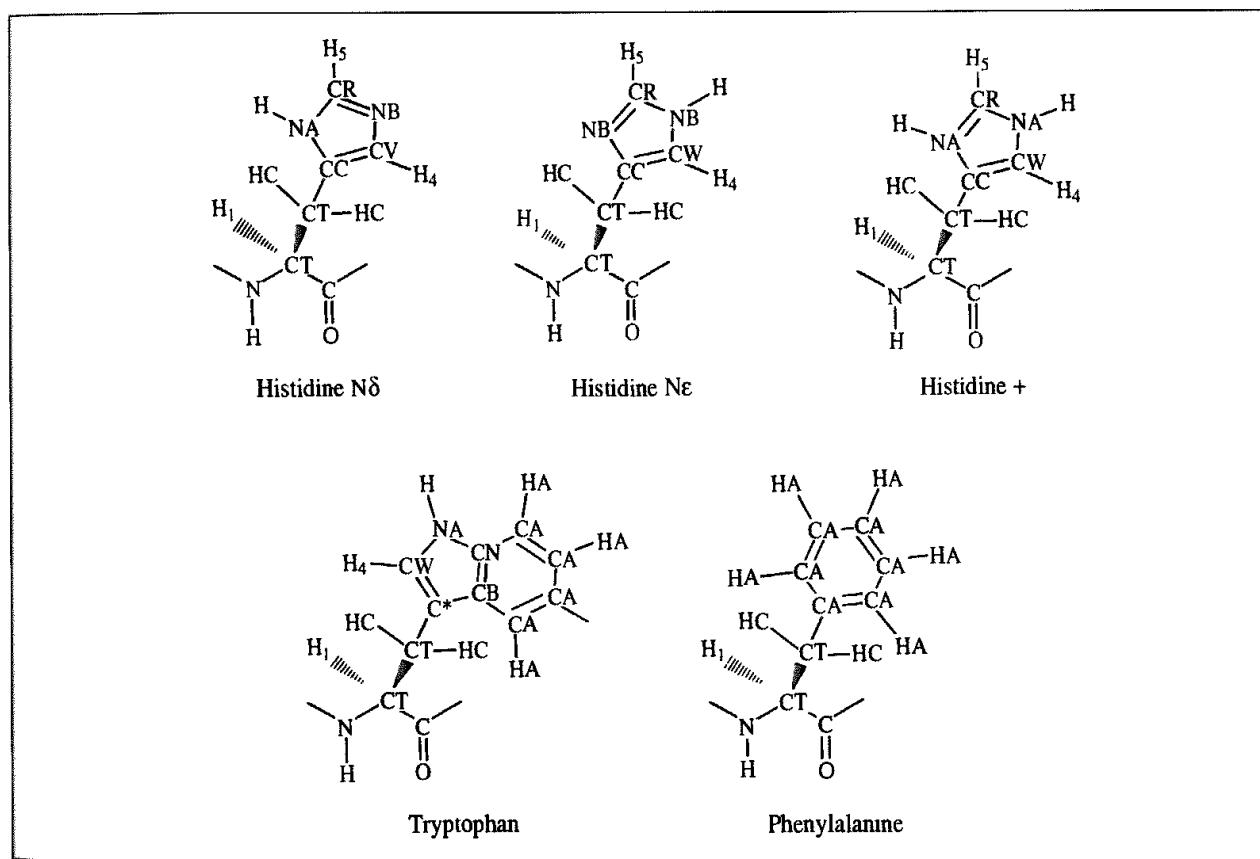


Fig. 4.3 AMBER atom types for the amino acids histidine, tryptophan and phenylalanine. There are three possible protonation states of histidine.

an atom type that is different from the carbon atom in an isolated five-membered ring such as histidine, which in turn is different from the atom type of a carbon atom in a benzene ring. Indeed, the AMBER force field uses different atom types for a histidine amino acid depending upon its protonation state (Figure 4.3). Other, more general, force fields would assign these atoms to the same generic 'sp² carbon' atom type. It is often found that force fields which are designed for modelling specific classes of molecule (such as proteins and nucleic acids, in the case of AMBER) use more specific atom types than force fields designed for general-purpose use.

We now discuss in some detail the individual contributions to a molecular mechanics force field, giving a selection of the various functional forms that are in common use. We shall then consider the important task of parametrisation, in which values for the many force constants are derived. Our discussion will be illuminated by examples chosen from contemporary force fields in widespread use and the MM2/MM3/MM4 and AMBER force fields in particular.

4.3 Bond Stretching

The potential energy curve for a typical bond has the form shown in Figure 4.4. Of the many functional forms used to model this curve, that suggested by Morse is particularly useful. The Morse potential has the form:

$$\nu(l) = D_e \{1 - \exp[-a(l - l_0)]\}^2 \quad (4.2)$$

D_e is the depth of the potential energy minimum and $a = \omega\sqrt{\mu/2D_e}$, where μ is the reduced mass and ω is the frequency of the bond vibration. ω is related to the stretching constant of the bond, k , by $\omega = \sqrt{k/\mu}$. l_0 is the reference value of the bond. The Morse potential is not usually used in molecular mechanics force fields. In part this is because it is not particularly amenable to efficient computation but also because it requires three parameters to be specified for each bond. Moreover, it is rare in molecular mechanics calculations for

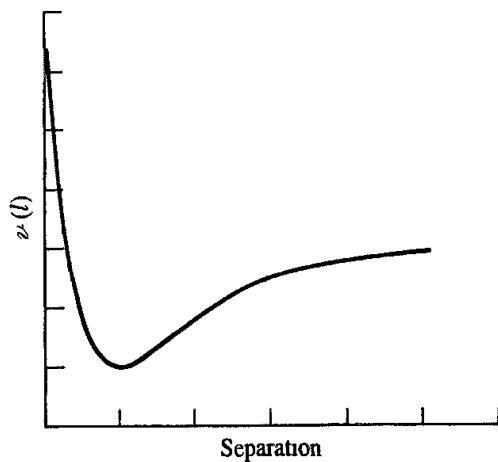


Fig. 4.4: Variation in bond energy with interatomic separation.

bonds to deviate significantly from their equilibrium values; the Morse curve describes a wide range of behaviour from the strong equilibrium behaviour to dissociation. Consequently, simpler expressions are often used. The most elementary approach is to use a Hooke's law formula in which the energy varies with the square of the displacement from the reference bond length l_0 :

$$\nu(l) = \frac{k}{2}(l - l_0)^2 \quad (4.3)$$

The astute reader will have noticed our use of the term 'reference bond length' (sometimes called the 'natural bond length') for the parameter l_0 . This parameter is commonly called the 'equilibrium' bond length, but to do so can be misleading. The reference bond length is the value that the bond adopts when all other terms in the force field are set to zero. The equilibrium bond length, by contrast, is the value that is adopted in a minimum energy structure, when all other terms in the force field contribute. The complex interplay between the various components in the force field means that the bond may well deviate slightly from its reference value in order to compensate for other contributions to the energy. It is also important to recognise that 'real' molecules undergo vibrational motion (even at absolute zero, there is a zero-point energy due to vibrational motion). A true bond-stretching potential is not harmonic but has a shape similar to that in Figure 4.4, which means that the 'average' length of the bond in a vibrating molecule will deviate from the equilibrium value for the hypothetical motionless state. The effects are usually small, but they are significant if one wishes to predict bond lengths to thousandths of an ångström. When comparing the results of calculations with experimental data, one must also remember that different experimental techniques measure different 'equilibrium' values, especially when the experiments are performed at different temperatures. The errors in experimentally determined bond lengths can be quite large; for example, libration of a molecule in a crystal means that the bond lengths determined by X-ray methods at room temperature may have errors as large as 0.015 Å. MM2 was parametrised to fit the values obtained by electron diffraction, which give the mean distances between atoms averaged over the vibrational motion at room temperature.

The forces between bonded atoms are very strong and considerable energy is required to cause a bond to deviate significantly from its equilibrium value. This is reflected in the magnitude of the force constants for bond stretching; some typical values from the MM2 force field are shown in Table 4.1, where it can be seen that those bonds one would

Bond	l_0 (Å)	k (kcal mol ⁻¹ Å ⁻²)
Csp ³ –Csp ³	1.523	317
Csp ³ –Csp ²	1.497	317
Csp ² =Csp ²	1.337	690
Csp ² =O	1.208	777
Csp ³ –Nsp ³	1.438	367
C–N (amide)	1.345	719

Table 4.1 Force constants and reference bond lengths for selected bonds [Allinger 1977]

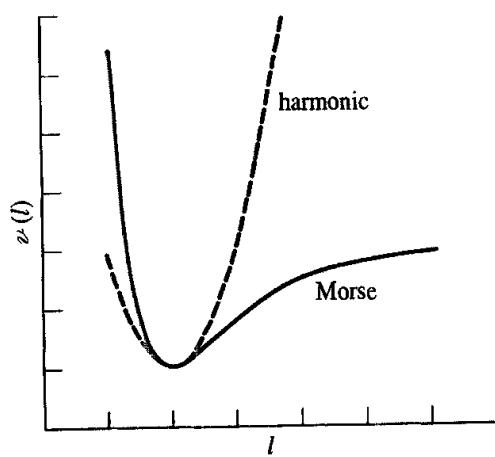


Fig. 4.5: Comparison of the simple harmonic potential (Hooke's law) with the Morse curve

intuitively expect to be stronger have large force constants (contrast C–C with C=C and N≡N). A deviation of just 0.2 Å from the reference value l_0 with a force constant of 300 kcal mol⁻¹ Å⁻² would cause the energy of the system to rise by 12 kcal/mol.

The Hooke's law functional form is a reasonable approximation to the shape of the potential energy curve at the bottom of the potential well, at distances that correspond to bonding in ground-state molecules. It is less accurate away from equilibrium (Figure 4.5). To model the Morse curve more accurately, cubic and higher terms can be included and the bond-stretching potential can be written as follows:

$$v(l) = \frac{k}{2}(l - l_0)^2[1 - k'(l - l_0) - k''(l - l_0)^2 - k'''(l - l_0)^3 \dots] \quad (4.4)$$

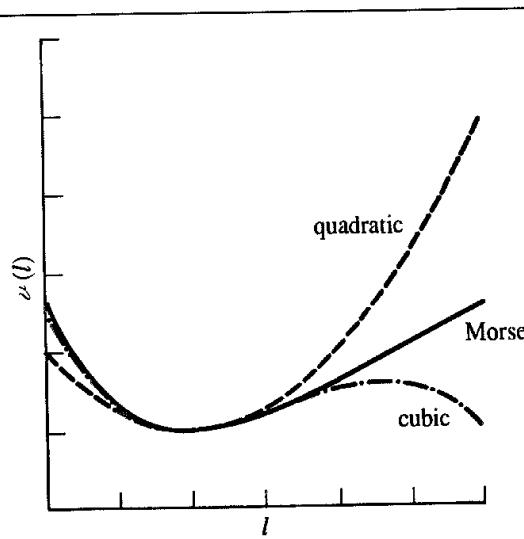


Fig. 4.6: A cubic bond-stretching potential passes through a maximum but gives a better approximation to the Morse curve close to the equilibrium structure than the quadratic form

An undesirable side-effect of an expansion that includes just a quadratic and a cubic term (as is employed in MM2) is that, far from the reference value, the cubic function passes through a maximum. This can lead to a catastrophic lengthening of bonds (Figure 4.6). One way to accommodate this problem is to use the cubic contribution only when the structure is sufficiently close to its equilibrium geometry and is well inside the ‘true’ potential well. MM3 also includes a quartic term; this eliminates the inversion problem and leads to an even better description of the Morse curve.

4.4 Angle Bending

The deviation of angles from their reference values is also frequently described using a Hooke’s law or harmonic potential:

$$\nu(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \quad (4.5)$$

The contribution of each angle is characterised by a force constant and a reference value. Rather less energy is required to distort an angle away from equilibrium than to stretch or compress a bond, and the force constants are proportionately smaller, as can be observed in Table 4.2.

Angle	θ_0	k (kcal mol ⁻¹ deg ⁻¹)
Csp ³ —Csp ³ —Csp ³	109.47	0.0099
Csp ³ —Csp ³ —H	109.47	0.0079
H—Csp ³ —H	109.47	0.0070
Csp ³ —Csp ² —Csp ³	117.2	0.0099
Csp ³ —Csp ² —Csp ²	121.4	0.0121
Csp ³ —Csp ² —O	122.5	0.0101

Table 4.2 Force constants and reference angles for selected angles
[Allinger 1977].

As with the bond-stretching terms, the accuracy of the force field can be improved by the incorporation of higher-order terms. MM2 contains a quartic term in addition to the quadratic term. Higher-order terms have also been included to treat certain pathological cases such as very highly strained molecules. The general form of the angle-bending term then becomes:

$$\nu(\theta) = \frac{k}{2}(\theta - \theta_0)^2[1 - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 - k'''(\theta - \theta_0)^3 \dots] \quad (4.6)$$

4.5 Torsional Terms

The bond-stretching and angle-bending terms are often regarded as ‘hard’ degrees of freedom, in that quite substantial energies are required to cause significant deformations from

their reference values. Most of the variation in structure and relative energies is due to the complex interplay between the torsional and non-bonded contributions.

The existence of barriers to rotation about chemical bonds is fundamental to understanding the structural properties of molecules and conformational analysis. The three minimum-energy staggered conformations and three maximum-energy eclipsed structures of ethane are a classic example of the way in which the energy changes with a bond rotation. Quantum mechanical calculations suggest that this barrier to rotation can be considered to arise from antibonding interactions between the hydrogen atoms on opposite ends of the molecule; the antibonding interactions are minimised when the conformation is staggered and are at a maximum when the conformation is eclipsed. Many force fields are used for modelling flexible molecules where the major changes in conformation are due to rotations about bonds; in order to simulate this it is essential that the force field properly represents the energy profiles of such changes.

Not all molecular mechanics force fields use torsional potentials; it may be possible to rely upon non-bonded interactions between the atoms at the end of each torsion angle (the 1,4 atoms) to achieve the desired energy profile. However, most force fields for 'organic' molecules do use explicit torsional potentials with a contribution from each bonded quartet of atoms A–B–C–D in the system. Thus there would be nine individual torsional terms for ethane and 24 for benzene ($6 \times C-C-C-C$, $12 \times C-C-C-H$ and $6 \times H-C-C-H$). Torsional potentials are almost always expressed as a cosine series expansion. One functional form is:

$$\nu(\omega) = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (4.7)$$

ω is the torsion angle.

An alternative but equivalent expression is:

$$\nu(\omega) = \sum_{n=0}^N C_n \cos(\omega)^n \quad (4.8)$$

V_n in Equation (4.7) is often referred to as the 'barrier' height, but to do so is misleading, obviously so when more than one term is present in the expansion. Moreover, other terms in the force field equation contribute to the barrier height as a bond is rotated, especially the non-bonded interactions between the 1,4 atoms. The value of V_n does, however, give a qualitative indication of the relative barriers to rotation; for example, V_n for an amide bond will be larger than for a bond between two sp^3 carbon atoms. n in Equation (4.7) is the *multiplicity*; its value gives the number of minimum points in the function as the bond is rotated through 360° . γ (the phase factor) determines where the torsion angle passes through its minimum value. For example, the energy profile for rotation about the single bond between two sp^3 carbon atoms could be represented by a single torsional term with $n = 3$ and $\gamma = 0^\circ$. This would give a threefold rotational profile with minima at torsion angles of $+60^\circ$, -60° and 180° and maxima at $\pm 120^\circ$ and 0° . A double bond between two sp^2 carbon atoms would have $n = 2$ and $\gamma = 180^\circ$, giving minima at 0° and 180° . The value of V_n would also be significantly larger for the double bond than for the single

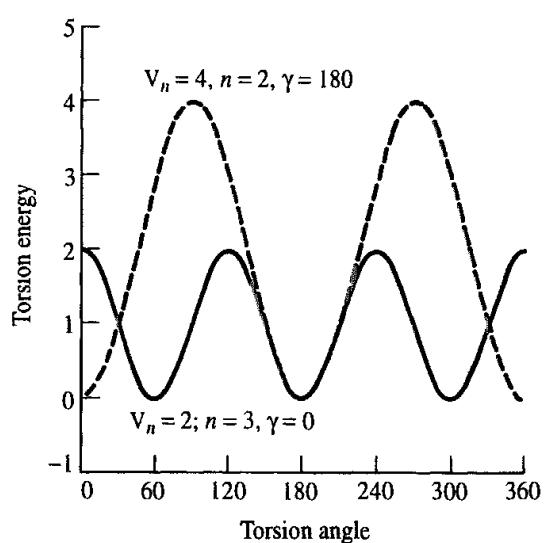


Fig 4.7. Torsional potential varies as shown for different values of V_n , n and γ .

bond. The effects of varying V_n , n and γ are illustrated in Figure 4.7 for commonly occurring torsional potentials.

Many of the torsional terms in the AMBER force field contain just one term from the cosine series expansion, but for some bonds it was found necessary to include more than one term. For example, to correctly model the tendency of O—C—C—O bonds to adopt a *gauche* conformation, a torsional potential with two terms was used for the O—C—C—O contribution:

$$\nu(\omega_{\text{C-O-O-C}}) = 0.25(1 + \cos 3\omega) + 0.25(1 + \cos 2\omega) \quad (4.9)$$

The torsional energy for a $\text{OCH}_2\text{—CH}_2\text{O}$ fragment (found in the sugars in DNA) varies with the torsion angle ω as shown in Figure 4.8. Another feature of the AMBER force field is its use of general torsional parameters. The energy profile for rotation about a bond that is described by a general torsional potential depends solely upon the atom types of the two

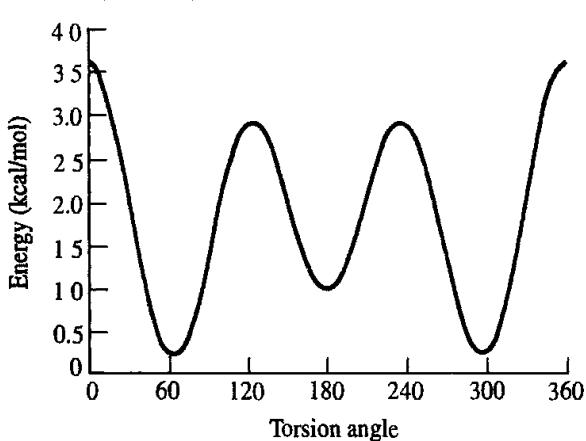


Fig 4.8: Variation in torsional energy (AMBER force field) with O—C—C—O torsion angle (ω) for $\text{OCH}_2\text{—CH}_2\text{O}$ fragment. The minimum energy conformations arise for $\omega = 60^\circ$ and 300°

atoms that comprise the central bond and not upon the atom types of the terminal atoms. For example, all torsion angles in which the central bond is between two sp^3 -hybridised carbon atoms (e.g. H–C–C–H, C–C–C–C, H–C–C–C) are assigned the same torsional parameters, unless the torsion is a special case such as O–C–C–O. In its treatment of the torsional contribution, AMBER takes a position intermediate between those force fields which only ever use a single term in the torsional expansion and those which consistently use more terms for all torsions. MM2 falls into the latter category; it uses three terms in the expansion:

$$\nu(\omega) = \frac{V_1}{2} (1 + \cos \omega) + \frac{V_2}{2} (1 - \cos 2\omega) + \frac{V_3}{2} (1 + \cos 3\omega) \quad (4.10)$$

A physical interpretation has been ascribed to each of the three terms in the MM2 torsional expansion from an analysis of *ab initio* calculations on simple fluorinated hydrocarbons. The first, onefold term corresponds to interactions between bond dipoles, which are due to differences in electronegativity between bonded atoms. The twofold term is due to the effects of hyperconjugation (in alkanes) and conjugation effects (in alkenes), which provide ‘double bond’ character to the bond. The threefold term corresponds to steric interactions between the 1,4 atoms. It was found that the additional terms in the torsional potential were especially important for systems containing heteroatoms, such as the halogenated hydrocarbons and molecules containing CCOC and CCNC fragments.

With careful parametrisation a force field which uses more than one term in the torsional expansion will be more successful than a force field that uses only a single term (and this is borne out by the MM2 force field). The major drawback is that many parameters are required to model even a modest range of molecules.

4.6 Improper Torsions and Out-of-plane Bending Motions

Let us consider how cyclobutanone would be modelled using a force field containing just standard bond-stretching and angle-bending terms of the type in Equation (4.1). The equilibrium structure obtained with such a force field would have the oxygen atom located out of the plane formed by the adjoining carbon atom and the two carbon atoms bonded to it, as shown in Figure 4.9. In this structure, the angles to the oxygen adopt values close to the reference value of 120° . Experimentally, it is found that the oxygen atom remains in the plane of the cyclobutane ring, even though the C–C=O angles are large (133°). This is because the π -bonding energy, which is maximised in the coplanar arrangement, would be much reduced if the oxygen were bent out of the plane. To achieve the desired geometry it is necessary to incorporate an additional term (or terms) in the force field that keeps the sp^2 carbon and the three atoms bonded to it in the same plane. The simplest way to achieve this is to use an *out-of-plane* bending term.

There are several ways in which out-of-plane bending terms can be incorporated into a force field. One approach is to treat the four atoms as an ‘improper’ torsion angle (i.e. a torsion angle in which the four atoms are not bonded in the sequence 1–2–3–4). One way to define an improper torsion for cyclobutane would involve the atoms 1–5–3–2 in Figure 4.9.

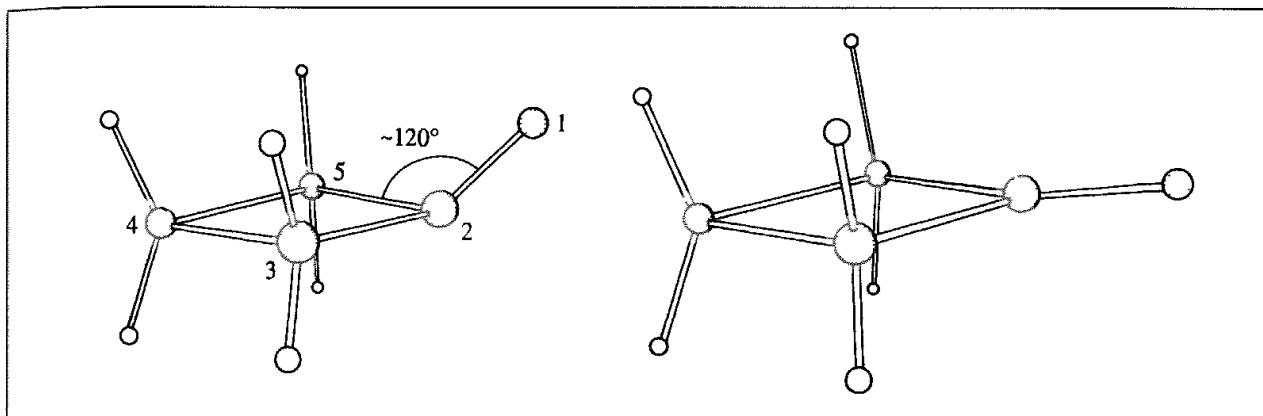


Fig. 4.9. Without an out-of-plane term, the oxygen atom in cyclobutane is predicted to lie out of the plane of the ring (left) rather than in the plane.

A torsional potential of the following form is then used to maintain the improper torsion angle at 0° or 180° :

$$\nu(\omega) = k(1 - \cos 2\omega) \quad (4.11)$$

Various other ways to incorporate the out-of-plane bending contribution are possible. For example, one definition that is closer to the notion of an 'out-of-plane bend' involves a calculation of the angle between a bond from the central atom and the plane defined by the central atom and the other two atoms (Figure 4.10). A value of 0° corresponds to all four atoms being coplanar. A third approach is to calculate the height of the central atom above a plane defined by the other three atoms (Figure 4.10). With these two definitions the deviation of the out-of-plane coordinate (be it an angle or a distance) can be modelled using a harmonic potential of the form

$$\nu(\theta) = \frac{k}{2}\theta^2; \quad \nu(h) = \frac{k}{2}h^2 \quad (4.12)$$

Of these three functional forms, the improper torsion definition is most widely used as it can then be easily included with the 'proper' torsional terms in the force field. However, the

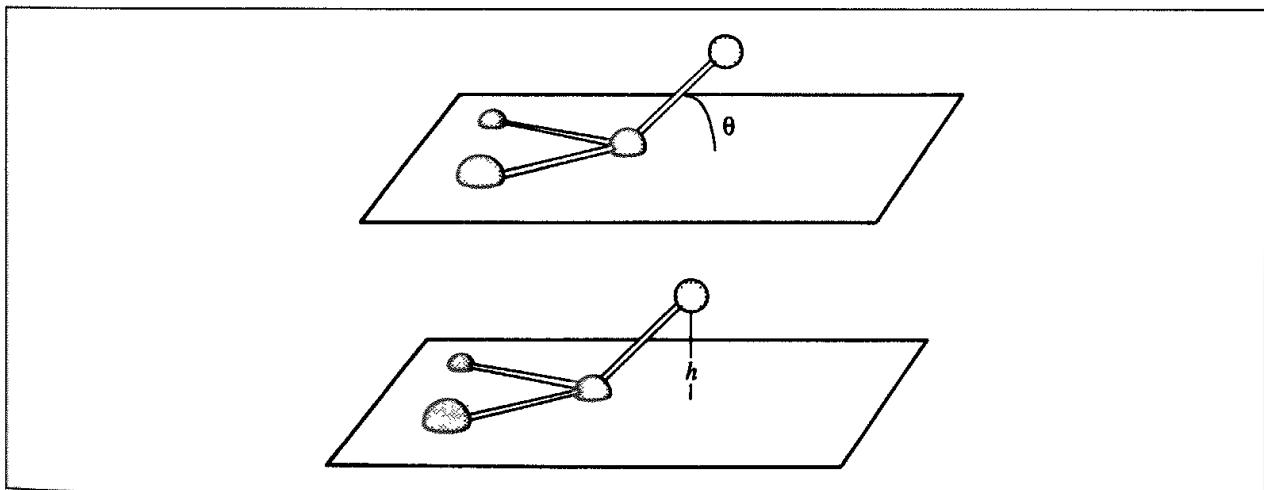


Fig. 4.10: Two ways to model the out-of-plane bending contributions.

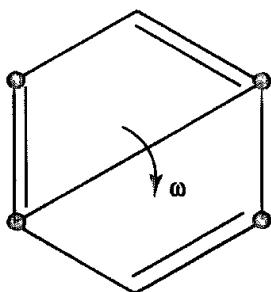


Fig. 4.11. Improper torsional terms can be used to keep a benzene ring planar

other two functional forms may be better ways to implement out-of-plane bending in the force field. Out-of-plane terms may also be used to achieve a particular geometry. For example, if it is desired to ensure that an aromatic ring such as benzene maintains an approximately planar structure then this can be achieved using a suitable set of out-of-plane bending terms involving atoms on opposite sides of the ring (Figure 4.11). Improper torsional terms are commonly used in the so-called united atom force fields to maintain stereochemistry at chiral centres (see Section 4.14). It is important to remember that out-of-plane terms may not always be necessary, and that to include such terms may have a deleterious effect on the performance of the force field. Vibrational frequencies in particular are often rather sensitive to the presence of out-of-plane terms.

4.7 Cross Terms: Class 1, 2 and 3 Force Fields

The presence of *cross terms* in a force field reflects coupling between the internal coordinates. For example, as a bond angle is decreased it is found that the adjacent bonds stretch to reduce the interaction between the 1,3 atoms, as illustrated in Figure 4.12. Cross terms were found to be important in force fields designed to predict vibrational spectra that were the forerunners of molecular mechanics force fields, and so it is not surprising that cross terms must often be included in a molecular mechanics force field to achieve optimal performance. One should in principle include cross terms between all contributions to a force field. However, only a few cross terms are generally found to be necessary in order to reproduce structural properties accurately; more may be needed to reproduce other properties such as vibrational frequencies, which are more sensitive to the presence of such terms. In general, any interactions involving motions that are far apart in a molecule

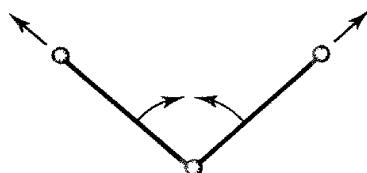


Fig. 4.12. Coupling between the stretching of the bonds as an angle closes.

can usually be set to zero. Most cross terms are functions of two internal coordinates, such as stretch–stretch, stretch–bend and stretch–torsion terms, but cross terms involving more than two internal coordinates such as the bend–bend–torsion have also been used. Various functional forms are possible for the cross terms. For example, the stretch–stretch cross term between two bonds 1 and 2 can be modelled as:

$$\nu(l_1, l_2) = \frac{k_{l_1, l_2}}{2} [(l_1 - l_{1,0})(l_2 - l_{2,0})] \quad (4.13)$$

The stretching of the two bonds adjoining an angle could be modelled using an equation of the following form (as in MM2, MM3 and MM4):

$$\nu(l_1, l_2, \theta) = \frac{k_{l_1, l_2, \theta}}{2} [(l_1 - l_{1,0}) + (l_2 - l_{2,0})](\theta - \theta_0) \quad (4.14)$$

In a *Urey–Bradley* force field, angle bending is achieved using 1,3 non-bonded interactions rather than an explicit angle-bending potential. The stretch-bond term in such a force field would be modelled by a harmonic function of the distance between the 1,3 atoms:

$$\nu(r_{1,3}) = \frac{k_{r_{1,3}}}{2} (r_{1,3} - r_{1,3}^0)^2 \quad (4.15)$$

A stretch-torsion cross term can be used to model the stretching of a bond that occurs in an eclipsed conformation. Two possible functional forms are:

$$\nu(l, \omega) = k(l - l_0) \cos n\omega \quad (4.16)$$

$$\nu(l, \omega) = k(l - l_0)[1 + \cos n\omega] \quad (4.17)$$

n is the periodicity of the rotation about the bond ($n = 3$ for sp^3 - sp^3 bonds).

Torsion-bend and torsion-bend-bend terms may also be included; the latter, for example, would couple two angles A–B–C and B–C–D to a torsion angle A–B–C–D. Maple, Dinur and Hagler used quantum mechanics calculations to investigate which of the cross terms are most important and suggested that the stretch–stretch, stretch–bend, bend–bend, stretch–torsion and bend–bend–torsion were most important [Dinur and Hagler 1991] (schematically illustrated in Figure 4.13).

It has been suggested that the presence of cross terms (together with some other features) can provide a general way to classify force fields [Hwang *et al.* 1994]. A class I force field was considered one which is restricted to harmonic terms (e.g. for bond stretching and angle bending) and which does not have any cross terms. A class II force field would have anharmonic terms (e.g. through the use of Morse potentials or quartic terms) and explicit cross terms to account for the coupling between coordinates. The presence of these higher and cross terms would tend to improve the ability of the force field to predict the properties of more unusual systems (such as those which are highly strained) and also to enhance its ability to reproduce vibrational spectra. Another characteristic of a class II force field was that it could be used without modification to model the properties of isolated small molecules, condensed phases and macromolecular systems. It was subsequently suggested by Allinger [Allinger *et al.* 1996b] that a class III force field would also take account of chemical effects and other features such as electronegativity and hyperconjugation. A classic

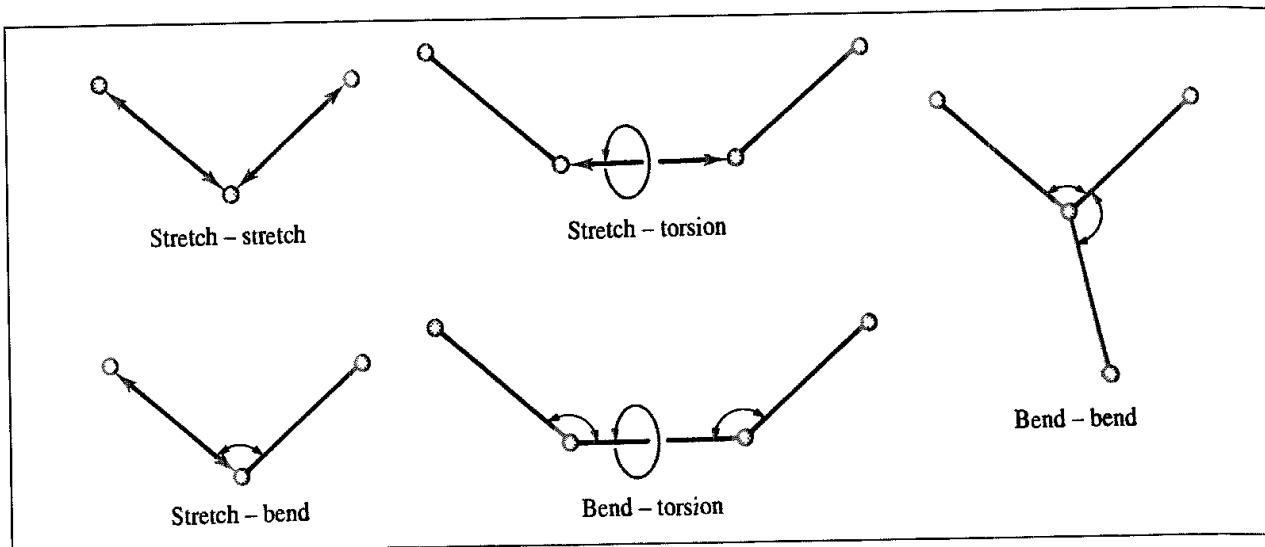


Fig. 4.13. Schematic illustration of the cross terms believed to be most important in force fields. (Adapted from Dinur U and A T Hagler 1991. New Approaches to Empirical Force Fields In Reviews in Computational Chemistry, Lipkowitz K B and D B Boyd (Editors) New York, VCH Publishers, pp 99–164)

example of the latter effect (hyperconjugation) is the change in the length of the C–H bond in acetaldehyde with rotation about the C–C bond. When the C–H bond is perpendicular to the plane of the carbonyl group there is maximum overlap between the σ orbital of the C–H bond and the π^* orbital of the carbonyl carbon. Donation of electron density from the C–H bond to this π^* orbital is accompanied by a lengthening of the bond and a greater contribution from the charged resonance structure (Figure 4.14). When the bond to the hydrogen atom is in the plane the overlap is minimal. *Ab initio* calculations suggested that the bond length changed by 0.006 Å between the two forms. This effect was incorporated within MM4 by a term of the following form:

$$\Delta l = k(1 - \cos 2\omega) \quad (4.18)$$

This is a kind of torsion–stretch cross term but different from the one where the central bond changes with torsion angle. There has been some considerable debate about the existence and origin of the hyperconjugative effects, but low-temperature X-ray crystallographic experiments on appropriate compounds together with *ab initio* calculations certainly reveal a detectable effect.

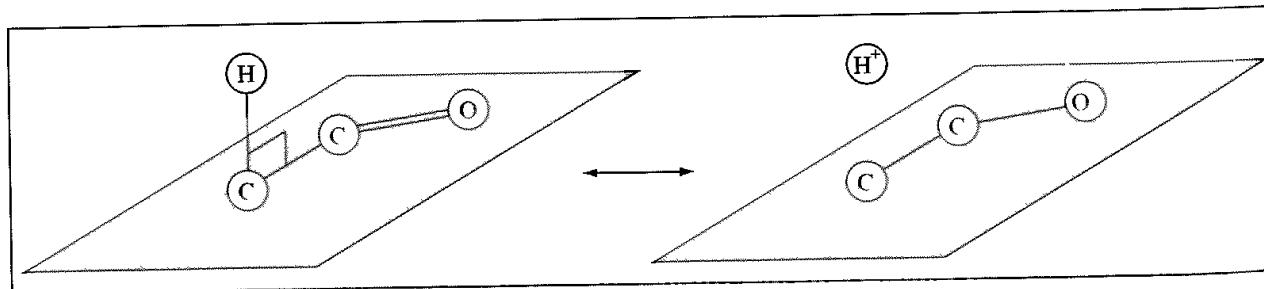


Fig. 4.14. Valence bond representation of the hyperconjugation effect which leads to a lengthening of the C–H bond in acetaldehyde

4.8 Introduction to Non-bonded Interactions

Independent molecules and atoms interact through non-bonded forces, which also play an important role in determining the structure of individual molecular species. The non-bonded interactions do not depend upon a specific bonding relationship between atoms. They are ‘through-space’ interactions and are usually modelled as a function of some inverse power of the distance. The non-bonded terms in a force field are usually considered in two groups, one comprising electrostatic interactions and the other van der Waals interactions.

4.9 Electrostatic Interactions

4.9.1 The Central Multipole Expansion

Electronegative elements attract electrons more than less electronegative elements, giving rise to an unequal distribution of charge in a molecule. This charge distribution can be represented in a number of ways, one common approach being an arrangement of fractional point charges throughout the molecule. These charges are designed to reproduce the electrostatic properties of the molecule. If the charges are restricted to the nuclear centres they are often referred to as *partial atomic charges* or *net atomic charges*. The electrostatic interaction between two molecules (or between different parts of the same molecule) is then calculated as a sum of interactions between pairs of point charges, using Coulomb’s law:

$$\mathcal{V} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (4.19)$$

N_A and N_B are the numbers of point charges in the two molecules. This approach to the representation and calculation of electrostatic interactions will be considered in more detail in Section 4.9.2. First, we shall consider an alternative approach to the calculation of electrostatic interactions which treats a molecule as a single entity and is (in principle at least) capable of providing a very efficient way to calculate electrostatic intermolecular interactions. This is the *central multipole expansion*, which is based upon the electric moments or multipoles: the charge, dipole, quadrupole, octopole, and so on introduced in Section 2.7.3. These moments are usually represented by the following symbols: q (charge), μ (dipole), Θ (quadrupole) and Φ (octopole). We are often interested in the lowest non-zero electric moment. Thus species such as Na^+ , Cl^- , NH_4^+ or CH_3CO_2^- have the charge as their lowest non-zero moment. For many uncharged molecules the dipole is the lowest non-zero moment. Molecules such as N_2 and CO_2 have the quadrupole as their lowest non-zero moment. The lowest non-zero moment for methane and tetrafluoromethane is the octopole. Each of these multipole moments can be represented by an appropriate distribution of charges. Thus a dipole can be represented using two charges placed an appropriate distance apart. A quadrupole can be represented using four charges and an octopole by eight charges. A complete description of the charge distribution around a molecule requires all of the non-zero electric moments to be specified. For some molecules,

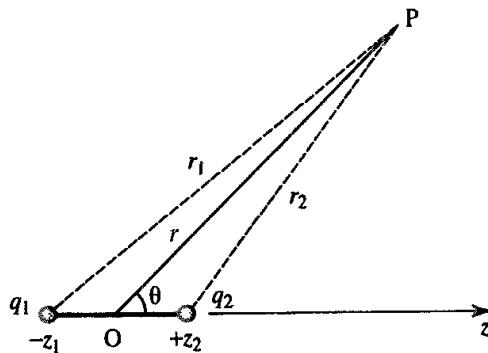


Fig. 4.15: The electrostatic potential due to two point charges.

the lowest non-zero moment may not be the most significant and it may therefore be unwise to ignore the higher-order terms in the expansion without first checking their values.

To illustrate how the multipolar expansion is related to a distribution of charges in a system, let us consider the simple case of a molecule with two charges q_1 and q_2 , positioned at $-z_1$ and z_2 , respectively (Figure 4.15). The electrostatic potential at point P (a distance r from the origin, r_1 from charge q_1 and r_2 from charge q_2) is then given by:

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{r_1} + \frac{q_2}{r_2} \right) \quad (4.20)$$

By applying the cosine rule this can be written as follows (see Figure 4.15):

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{\sqrt{r^2 + z_1^2 + 2rz_1 \cos\theta}} + \frac{q_2}{\sqrt{r^2 + z_2^2 - 2rz_2 \cos\theta}} \right) \quad (4.21)$$

If $r \gg z_1$ and $r \gg z_2$ then this expression can be expanded as follows:

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1 + q_2}{r} + \frac{(q_2 z_2 - q_1 z_1) \cos\theta}{r^2} + \frac{(q_1 z_1^2 + q_2 z_2^2)(3 \cos^2\theta - 1)}{2r^3} + \dots \right) \quad (4.22)$$

We can now associate the appropriate terms in the expansion with the various electric moments:

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{r} + \frac{\mu \cos\theta}{r^2} + \frac{\Theta(3 \cos^2\theta - 1)}{2r^3} + \dots \right) \quad (4.23)$$

Thus $(q_1 + q_2)$ is the charge; $(q_2 z_2 - q_1 z_1)$ is the dipole; $(q_1 z_1^2 + q_2 z_2^2)$ is the quadrupole, and so on. One interesting feature about a charge distribution is that only the first non-zero moment is independent of the choice of origin. Thus, if a molecule is electrically neutral (i.e. $q_1 + q_2 = 0$) then its dipole moment is independent of the choice of origin. This can be demonstrated for our two-charge system as follows. If the position of the origin is now moved to a point $-z'$, then the dipole moment relative to this new origin is given by:

$$\mu' = q_2(z_2 + z') - q_1(z_1 - z') = \mu + qz' \quad (4.24)$$

Only if the total charge on the system (q) equals zero will the dipole moment be unchanged. Similar arguments can be used to show that if both the charge and the dipole moment are zero then the quadrupole moment is independent of the choice of origin. For convenience, the origin is often taken to be the centre of mass of the charge distribution.

The electric moments are examples of *tensor properties*: the charge is a rank 0 tensor (which is the same as a scalar quantity); the dipole is a rank 1 tensor (which is the same as a vector, with three components along the x , y and z axes); the quadrupole is a rank 2 tensor with nine components, which can be represented as a 3×3 matrix. In general, a tensor of rank n has 3^n components.

For a distribution of charges (one not restricted to lie along one of the Cartesian axes), the dipole moment is given by:

$$\mu = \sum q_i r_i \quad (4.25)$$

The components of the dipole moment along the x , y and z axes are $\sum q_i x_i$, $\sum q_i y_i$ and $\sum q_i z_i$. The analogous way to define the quadrupole moment is as follows:

$$\Theta = \begin{pmatrix} \sum q_i x_i^2 & \sum q_i x_i y_i & \sum q_i x_i z_i \\ \sum q_i y_i x_i & \sum q_i y_i^2 & \sum q_i y_i z_i \\ \sum q_i z_i x_i & \sum q_i z_i y_i & \sum q_i z_i^2 \end{pmatrix} \quad (4.26)$$

This definition of the quadrupole is obviously dependent upon the orientation of the charge distribution within the coordinate frame. Transformation of the axes can lead to alternative definitions that may be more informative. Thus the quadrupole moment is commonly defined as follows:

$$\Theta = \frac{1}{2} \begin{pmatrix} \sum_i q_i (3x_i^2 - r_i^2) & 3 \sum_i q_i x_i y_i & 3 \sum_i q_i x_i z_i \\ 3 \sum_i q_i x_i z_i & \sum_i q_i (3y_i^2 - r_i^2) & 3 \sum_i q_i y_i z_i \\ 3 \sum_i q_i x_i z_i & 3 \sum_i q_i y_i z_i & \sum_i q_i (3z_i^2 - r_i^2) \end{pmatrix} \quad (4.27)$$

In Equation (4.27) $r_i^2 = x_i^2 + y_i^2 + z_i^2$. This definition enables one to assess the deviation from spherical symmetry as a spherically symmetric charge distribution will have

$$\sum_i q_i x_i^2 = \sum_i q_i y_i^2 = \sum_i q_i z_i^2 = \frac{1}{3} \sum_i q_i r_i^2 \quad (4.28)$$

and so the diagonal elements of the tensor will be zero. Quadrupoles are also reported in terms of the *principal axes*; these are three mutually perpendicular axes α , β and γ , which are linear combinations of x , y and z such that the quadrupole tensor is diagonal (i.e. off-diagonal elements are zero):

$$\Theta = \begin{pmatrix} \Theta_{\alpha\alpha} & 0 & 0 \\ 0 & \Theta_{\beta\beta} & 0 \\ 0 & 0 & \Theta_{\gamma\gamma} \end{pmatrix} \quad (4.29)$$

Let us now consider the effect of placing another molecule with a linear charge distribution (charges q'_1 and q'_2) with its centre of mass at the point P. The relative orientation of the two

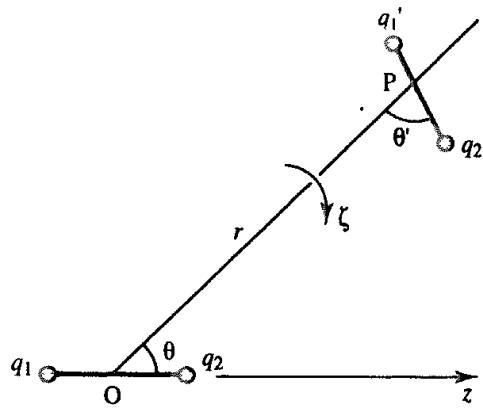


Fig. 4.16: The relative orientation of two dipoles

molecules can be described in terms of four parameters (the distance joining their centres of mass and three angles as shown in Figure 4.16). The electrostatic interaction between the two molecules is calculated by multiplying each charge by the potential at that point and adding the result for each charge. The following expression is the result [Buckingham 1959]:

$$\mathcal{V}(q, q') = \frac{1}{4\pi\epsilon_0} \left\{ \begin{aligned} & \frac{qq'}{r} \\ & + \frac{1}{r^2} (q\mu' \cos \theta + q'\mu \cos \theta') \\ & + \frac{\mu\mu'}{r^3} (2 \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \zeta) \\ & + \frac{1}{2r^3} [q\Theta'(3 \cos^2 \theta' - 1) + q'\Theta(3 \cos^2 \theta - 1)] \\ & + \frac{3}{2r^4} [\mu\Theta' \{ \cos \theta (3 \cos^2 \theta' - 1) + 2 \sin \theta \sin \theta' \cos \theta' \cos \zeta \} \\ & \quad + \mu' \Theta \{ \cos \theta' (3 \cos^2 \theta - 1) + 2 \sin \theta' \sin \theta \cos \theta \cos \zeta \}] \\ & + \frac{3\Theta\Theta'}{4r^5} [1 - 5 \cos^2 \theta - 5 \cos^2 \theta' + 17 \cos^2 \theta \cos^2 \theta' \\ & \quad + 2 \sin^2 \theta \sin^2 \theta' \cos^2 \zeta + 16 \sin \theta \sin \theta' \cos \theta \cos \theta' \cos \zeta] \\ & + \dots \end{aligned} \right\} \quad (4.30)$$

The energy of interaction between two charge distributions is thus an infinite series that includes charge-charge, charge-dipole, dipole-dipole, charge-quadrupole, dipole-quadrupole interactions, quadrupole-quadrupole terms, and so on. These terms depend on different inverse powers of the separation r . If the molecules are neutral (i.e. $q = q' = 0$) then the leading term in the expansion is that due to the dipole-dipole interaction, which varies as r^{-3} . This is a key result, for the range of the dipole-dipole interaction (r^{-3}) is much less than that of the Coulomb interaction (r^{-1}), Figure 4.17. This will be important in later chapters, where we shall collect atoms together into neutral groups. The electrostatic interaction

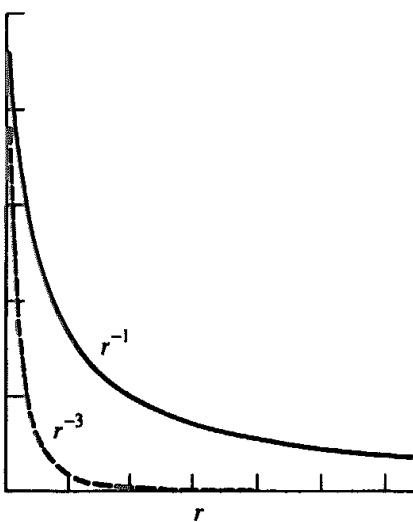


Fig. 4.17 The charge-charge energy decays much more slowly (αr^{-1}) than the dipole-dipole energy (αr^{-3})

between these groups then decays as r^{-3} rather than the r^{-1} dependence of each individual charge-charge interaction. This can be seen in Figure 4.17, in which the functions r^{-1} and r^{-3} have been plotted as a function of distance. Even when the dipole-dipole interaction energy has fallen off almost to zero the charge-charge interaction energy is still significant. In general, the interaction energy between two multipoles of order n and m decreases as $r^{-(n+m+1)}$. It should be emphasised again that these expressions are only valid when the separation of the two molecules, r , is much larger than the internal dimensions of the molecules. The favourable arrangements for the various multipoles are shown in Figure 4.18.

A central multipole expansion therefore provides a way to calculate the electrostatic interaction between two molecules. The multipole moments can be obtained from the wavefunction and can therefore be calculated using quantum mechanics (see Section 2.7.3) or can be determined from experiment. One example of the use of a multipole expansion is

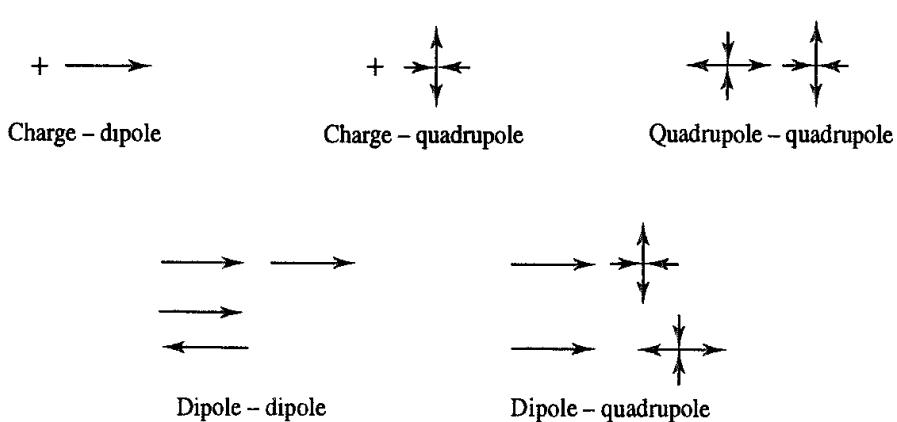


Fig. 4.18 The most favourable orientations of various multipoles (Figure adapted from Buckingham A D 1959 Molecular Quadrupole Moments. Quarterly Reviews of the Chemical Society 13:183–214.)

the benzene model of Claessens, Ferrario and Ryckaert [Claessens *et al.* 1983]. Benzene has no charge and no dipole moment, but it does have a sizeable quadrupole. The inclusion of the quadrupole was found to give clearly superior results in molecular dynamics simulations of the liquid state over models that lacked any electronic contribution.

The main advantage of the multipolar description for calculating the electrostatic interactions between molecules is its efficiency. For example, the charge-charge interaction energy between two benzene molecules would require 144 individual charge-charge interactions with a partial atomic charge model rather than the single quadrupole-quadrupole term. Unfortunately, the multipole expansion is not applicable when the molecules are separated by distances comparable with the molecular dimensions. The formal condition for convergence of the multipolar interaction energy is that the distance between two interacting molecules should be larger than the sum of the distances from the centre of each molecule to the furthest part of its charge distribution. If a sphere is constructed around each molecule, positioned on its centre of mass, with a radius that encompasses all of the charge distribution, then the multipole expansion for the interaction between two molecules will converge if these spheres do not intersect. Even if one requires the sphere to encompass just the nuclei in a molecule (i.e. ignoring the fact that the charge distribution around a molecule extends to infinity) there may still be problems. For example, the convergence sphere for a molecule such as butane would extend beyond the van der Waals radii in some directions, enabling other molecules to penetrate the convergence sphere, as illustrated in Figure 4.19. Another problem is that the multipolar expansion may be slow to converge. The multipolar expansion is often located at the centre of mass, but this may not be the best choice to achieve the most rapid convergence.

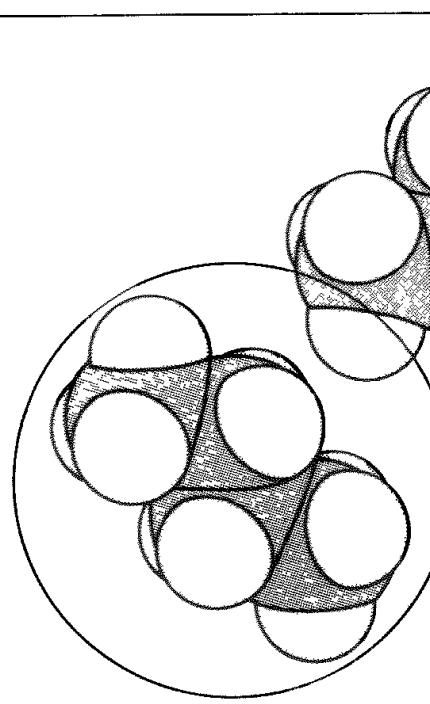


Fig. 4.19. The convergence sphere of the multipole expansion for a molecule such as butane may be penetrated by another molecule

There are other difficulties with the central multipole expansion. The multipole moments are properties of the entire molecule and so cannot be used to determine intramolecular interactions. The central multipole model thus tends to be restricted to calculations involving small molecules that are kept fixed in conformation during the calculation, and where the interactions between molecules act at their centres of mass. It can be a complicated procedure to calculate the forces acting on a molecule with a multipole model. The interaction between multipoles of zero order (i.e. charges) gives rise to a simple translational force. Multipoles of a higher order have directionality, and interactions between these produce a torque, or twisting force. Moreover, whereas the charge-charge forces are equal and opposite, the torque acting on molecule i due to another molecule j is not necessarily equal and opposite to the torque on molecule j due to molecule i .

4.9.2 Point-charge Electrostatic Models

We therefore return to the point-charge model for calculating electrostatic interactions. If sufficient point charges are used then all of the electric moments can be reproduced and the multipole interaction energy, Equation (4.30), is exactly equal to that calculated from the Coulomb summation, Equation (4.19).

An accurate representation of a molecule's electrostatic properties may require charges to be placed at locations other than at the atomic nuclei. A simple example of this is molecular nitrogen, which has a dipole moment of zero. The total charge on nitrogen is zero, and so an atomic partial charge model would put zero charge on each nucleus. However, nitrogen does have a quadrupole moment and this significantly affects its properties. The simplest way to model this is to place three partial charges along the bond: a charge of $-q$ at each nucleus and $+2q$ at the centre of mass. The quadrupole-quadrupole interaction between two nitrogen molecules can then be calculated by summing nine pairs of charge-charge interactions. The value of q can be calculated using the following relationship between the quadrupole moment and the partial charge:

$$\Theta = 2q(l/2)^2 \quad (4.31)$$

l is the bond length. The experimental quadrupole moment is consistent with a charge, q , of approximately $0.5e$. In fact, a better representation of the electrostatic potential around the nitrogen molecule is obtained using the five-charge model shown in Figure 4.20.

An alternative to the point charge model is to assign dipoles to the bonds in the molecule. The electrostatic energy is then given as a sum of dipole-dipole interaction energies. This approach (which is adopted in MM2/MM3/MM4) can be unwieldy for molecules that have a formal charge and which require charge-charge and charge-dipole terms to be included in the energy expression. Charged species are dealt with more naturally using the point charge model.

4.9.3 Calculating Partial Atomic Charges

Given the widespread use of the partial atomic charge model, it is important to consider how the charges are obtained. For simple species the atomic charges required to reproduce the

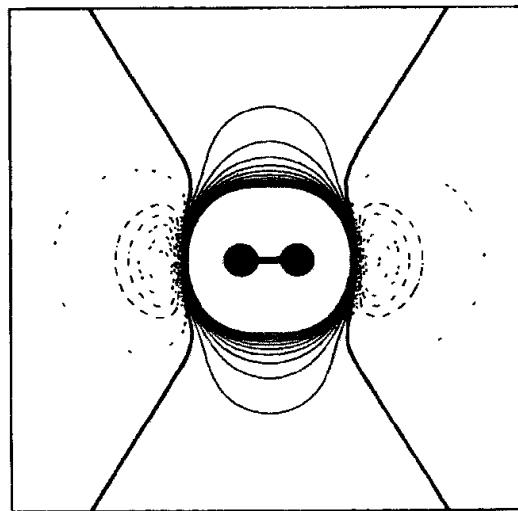
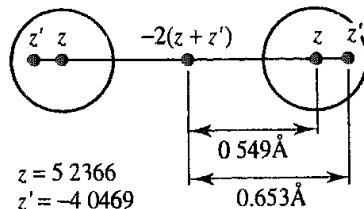
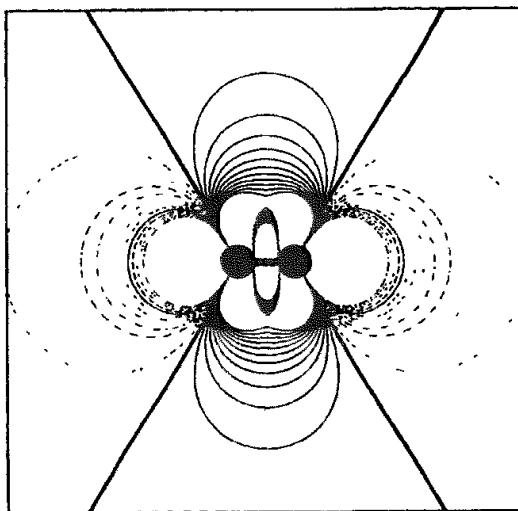
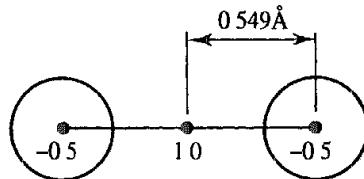
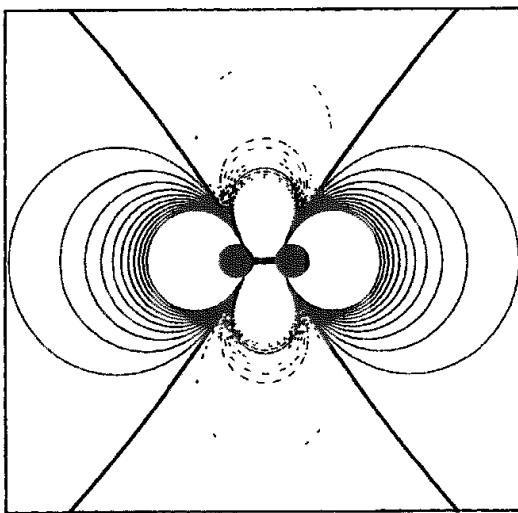


Fig. 4.20: Two charge models for N_2 with the electrostatic potentials that they generate. Also shown is the electrostatic potential calculated using ab initio quantum mechanics (6-31G^{*} basis set.) Negative contours are dashed and the zero contour is bold

electric moments can be calculated exactly if the geometry is known. For example, the experimentally determined dipole moment of HF (1.82 D) can be reproduced by placing equal but opposite charges of $0.413e$ on the two atomic nuclei (assuming a bond length of 0.917 \AA). The tetrahedral arrangement of the hydrogens about the carbon in methane means that each hydrogen atom has an identical charge equal to one quarter the charge on the carbon. The molecule is electrically neutral with zero dipole and quadrupole moments but a non-zero octopole moment, which can be reproduced using a hydrogen charge of approximately $0.14e$.

In some cases the atomic charges are chosen to reproduce thermodynamic properties calculated using a molecular dynamics or Monte Carlo simulation. A series of simulations is performed and the charge model is modified until satisfactory agreement with experiment is obtained. This approach can be quite powerful despite its apparent simplicity, but it is only really practical for small molecules or simple models.

The electrostatic properties of a molecule are a consequence of the distribution of the electrons and the nuclei and thus it is reasonable to assume that one should be able to obtain a set of partial atomic charges using quantum mechanics. Unfortunately, the partial atomic charge is not an experimentally observable quantity and cannot be unambiguously calculated from the wavefunction. This explains why numerous ways to determine partial atomic charges have been proposed, and why there is still considerable debate as to the 'best' method to derive them. Indirect comparisons of the various methods are possible, usually by calculating appropriate quantities from the charge model and then comparing the results with either experiment or quantum mechanics. For example, one might examine how well the charge model reproduces the experimental or quantum mechanical multipole moments or the electrostatic potential around the molecule.

We have already encountered in Section 2.7.5 the population analysis method for calculating partial atomic charges. Such sets of charges (commonly referred to as *Mulliken charges* when obtained from that particular partitioning scheme) are often considered to be inappropriate for accurately representing the interactions between molecules. This is because Mulliken charges are primarily dependent upon the constitution of the molecule – how the atoms are bonded together – rather than being designed to reproduce the properties that determine how molecules interact with each other, such as the electrostatic potential. The importance of the electrostatic potential in intermolecular interactions has resulted in much interest in schemes that calculate charges consistent with this particular property.

4.9.4 Charges Derived from the Molecular Electrostatic Potential

The electrostatic potential at a point is the force acting on a unit positive charge placed at that point. The nuclei give rise to a positive (i.e. repulsive) force, whereas the electrons give rise to a negative potential. The electrostatic potential is an observable quantity that can be determined from a wavefunction using Equations (2.222) and (2.223):

$$\phi(\mathbf{r}) = \phi_{\text{nucl}}(\mathbf{r}) + \phi_{\text{elec}}(\mathbf{r}) = \sum_{A=1}^M \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{d\mathbf{r}' \rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \quad (4.32)$$

The electrostatic potential is a continuous property and is not easily represented by an analytical function. Consequently, it is necessary to derive a discrete representation for use in numerical analysis. The objective is to derive the set of partial charges (usually partial atomic charges) that best reproduces the quantum mechanical electrostatic potential at a series of points surrounding the molecule. A solution to this problem was suggested by Cox and Williams [Cox and Williams 1981]. The electrostatic potential at each of the chosen points is calculated from the wavefunction. A least-squares fitting procedure is then employed to determine the set of partial atomic charges that best reproduces the electrostatic potential at the points, subject to the constraint that the sum of the charges should be equal to the net charge on the molecule. Symmetry conditions may also be imposed to ensure that the charges on symmetrically equivalent atoms are equal. It is also possible to require the atomic charges to reproduce other electrostatic properties of the molecules such as the dipole moment. The fitting procedure minimises the sum of squares of the differences in the electrostatic potential. Thus, if the electrostatic potential at a point is ϕ_i^0 and if the value from the charge model is ϕ_i^{calc} , then the objective is to minimise the following function:

$$R = \sum_{i=1}^{N_{\text{points}}} w_i (\phi_i^0 - \phi_i^{\text{calc}})^2 \quad (4.33)$$

N_{points} is the number of points and w_i is a weighting factor that enables different points to be given different degrees of ‘importance’ in the fitting process. One of the charges is dependent on the values of the others (because the sum must equal Z , the molecular charge). This N th charge has a value given by:

$$q_N = Z - \sum_{j=1}^{N-1} q_j \quad (4.34)$$

The electrostatic potential due to the charges q_j at the point i is given by Coulomb’s law:

$$\phi_i^{\text{calc}} = \sum_{j=1}^{N-1} \frac{q_j}{4\pi\epsilon_0 r_{ij}} + \frac{Z - \sum_{j=1}^{N-1} q_j}{4\pi\epsilon_0 r_{iN}} \quad (4.35)$$

r_{ij} is the distance from the charge j to the point i . At a minimum value of the error function, R , the first derivative is equal to zero with respect to all charges q_k :

$$\frac{\partial R}{\partial q_k} = -2 \sum_{i=1}^{N_{\text{points}}} w_i (\phi_i^0 - \phi_i^{\text{calc}}) \left(\frac{\partial \phi_i^{\text{calc}}}{\partial q_k} \right) = 0 \quad (4.36)$$

This equation can be written in the following form:

$$\sum_{i=1}^{N_{\text{points}}} w_i \left(\phi_i^0 - \frac{Z}{r_{iN}} \right) \left(\frac{1}{r_{ik}} - \frac{1}{r_{iN}} \right) = \sum_{j=1}^{N-1} \left[\sum_{i=1}^{N_{\text{points}}} w_i \left(\frac{1}{r_{ik}} - \frac{1}{r_{iN}} \right) \left(\frac{1}{r_{ij}} - \frac{1}{r_{iN}} \right) \right] \frac{q_j}{4\pi\epsilon_0} \quad (4.37)$$

When expressed in this way, then the set of equations can be recast as a matrix equation of the form $\mathbf{A}\mathbf{q} = \mathbf{a}$. The charges \mathbf{q} are then determined using standard matrix methods via $\mathbf{q} = \mathbf{A}^{-1}\mathbf{a}$.

The points i ($1, 2, \dots, N_{\text{points}}$) where the potential is fitted can be chosen in a variety of ways but should be taken from the region where it is most important to model intermolecular interactions correctly. This region is just beyond the van der Waals radii of the atoms involved. Cox and Williams selected points from a regular grid in a shell defined by two surfaces, one corresponding to the union of the van der Waals radii plus 1.2 \AA and the others approximately 1 \AA beyond that. The CHELP procedure of Chirlian and Franci [Chirlian and Franci 1987] uses spherical shells, 1 \AA apart, centred on each atom with points symmetrically distributed on the surface. Any points within the van der Waals radius of any atom in the system are discarded and the shells extend to 3 \AA from the van der Waals surface of the molecule. The CHELP method employs a Lagrange multiplier method to find the atomic charges, rather than an iterative least-squares procedure. This minimises the error function R (Equation (4.33)) subject to the constraint that the charges sum to the total molecular charge. Such an analysis yields a set of $N + 1$ equations in $N + 1$ unknowns and can be solved using standard matrix methods. The CHELPG algorithm of Breneman and Wiberg [Breneman and Wiberg 1990] combines the regular grid of points of Cox and Williams with the Lagrange multiplier method of Chirlian and Franci as the results from CHELP were found to change if the molecule was reoriented in the coordinate system. In CHELPG a cubic grid of points (spaced $0.3\text{--}0.8 \text{ \AA}$ apart) is used and all grid points that lie within the van der Waals radius of any atom are discarded, together with all points that lie further than 2.8 \AA away from any atom.

The algorithm of Singh and Kollman used to derive the charges in the 1984 AMBER force field uses points on a series of molecular surfaces, constructed using gradually increasing van der Waals radii for the atoms [Singh and Kollman 1984]. The points at which the potential was fitted were located on these shells. For the 1995 AMBER force field a modified version of this electrostatic potential method was employed (termed ‘restrained electrostatic potential fit’, or RESP [Bayly *et al.* 1993]). The RESP algorithm uses hyperbolic restraints on non-hydrogen atoms. These restraints have the effect of reducing the charges on some atoms, particularly buried carbon atoms, which can be assigned artificially high charges in standard electrostatic potential fitting methods. The RESP charges also vary less with the molecular conformation.

4.9.5 Deriving Charge Models for Large Systems

Molecular mechanics is used to model systems containing thousands of atoms such as polymers. How then can charges be derived for such species? Clearly one cannot routinely perform quantum mechanical calculations on a molecule with so many atoms and so it must be broken into fragments of a suitable size. In some cases the fragments might appear relatively easy to define; for example, many polymeric systems are constructed by connecting together chemically defined monomeric units. The atomic charges for each monomer should be obtained from calculations on suitable fragments that recreate the immediate local environment of the fragment in the larger molecule. For example, partial atomic charges for amino acids are often obtained from calculations on a ‘dipeptide’ fragment (see Figure 4.21), which is more akin to the environment within a protein than in an isolated amino acid.

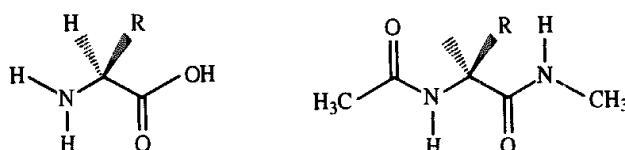


Fig 4.21 The charges used for calculations on proteins are best derived using a suitable fragment for each amino acid that reflects the environment within the protein (right), rather than the isolated amino acid (left)

The charge sets obtained from electrostatic potential fitting can be highly dependent upon the basis set used to derive the wavefunction. Moreover, the charges do not always improve if a larger basis set is used. It is generally considered that the 6-31G* basis set gives reasonable results for calculations relevant to condensed phases. In many cases it is possible to scale the results of a calculation using a small basis set or even a lower level of theory (such as a semi-empirical calculation) to obtain results comparable with those of a high-level calculation. Of the various semi-empirical methods available, MNDO appears to give the best correspondence with the charges derived from *ab initio* calculations, and scaling factors have been determined by several research groups [Ferenczy *et al.* 1990; Luque *et al.* 1990; Bezler *et al.* 1990]. An additional complicating factor is that the charges obtained from electrostatic potential fitting will often depend upon the conformation for which the quantum mechanical calculation was performed [Williams 1990]. One solution is to perform a series of charge calculations for different conformations and then use a charge model in which each charge is weighted according to the relative population of that particular conformation as calculated from the Boltzmann distribution [Reynolds *et al.* 1992]. In a few charge models the charges vary continuously with the conformation [Rappé and Goddard 1991; Dinur and Hagler 1995].

4.9.6 Rapid Methods for Calculating Atomic Charges

Some methods calculate atomic charges solely from information about the atoms present in the molecule and the way in which the atoms are connected. The great advantage of such methods is that they are very fast and can be used to calculate the charge distributions for large numbers of molecules (e.g. in a database). We will consider the Gasteiger and Marsili method [Gasteiger and Marsili 1980] as an example.

The Gasteiger–Marsili approach uses the concept of the *partial equalisation of orbital electronegativity*. Electronegativity is a concept well known to chemists, being defined by Pauling as ‘the power of an atom to attract electrons to itself’. Mulliken subsequently defined the electronegativity of an atom A as the average of its ionisation potential I_A and its electron affinity E_A :

$$\chi_A = \frac{1}{2}(I_A + E_A) \quad (4.38)$$

As Mulliken pointed out, the ionisation potential and electron affinity are specific to a given valence state of an atom, and therefore the electronegativities of an atom’s valence states would not be expected to be the same. This idea can be extended to the concept of orbital

electronegativity, which is the electronegativity of a specific orbital in a given valence state. For example, an sp orbital has a higher electronegativity than an sp^3 orbital. The orbital electronegativity will also depend on the occupancy of the orbital; an empty orbital will be better able to attract an electron than an orbital with a single electron, which in turn will be better than an orbital with two electrons. The electronegativity of an orbital will also be affected by the charges in other orbitals. Gasteiger and Marsili assumed a polynomial relationship between the orbital electronegativity $\chi_{\mu A}$ of an orbital ϕ_μ in atom A and the charge Q_A on the atom A:

$$\chi_{\mu A} = a_\mu + b_{\mu A} Q_A + c_{\mu A} Q_A^2 \quad (4.39)$$

Values of the coefficients a , b and c were derived for common elements in their usual valence states (for example, for carbon there are different values for sp^3 , $sp^2\pi$ and $sp\pi^2$ valence states).

Electrons flow from the less electronegative elements to the more electronegative ones. This flow of electrons results in a positive charge on the less electronegative atoms and a negative charge on the more electronegative atoms, and as such the flow acts to equalise the electronegativities. Total equalisation of electronegativity does not, however, lead to chemically sensible results. This effect is modelled in the Gasteiger and Marsili approach by an iterative procedure, in which less and less charge is transferred between bonded atoms at each step. The electron charge transferred from an atom A to an atom B (where B is more electronegative than A) in iteration k is given by:

$$Q^{(k)} = \frac{\chi_B^{(k)} - \chi_A^{(k)}}{\chi_A^+} \alpha^k \quad (4.40)$$

In Equation (4.40), $Q^{(k)}$ is the charge (in electrons) transferred; $\chi_A^{(k)}$ and $\chi_B^{(k)}$ are the electronegativities of the atoms A and B; χ_A^+ is the electronegativity of the cation of the less electronegative atom and α is a damping factor which is raised to the power k . Gasteiger and Marsili set α to $\frac{1}{2}$. The charge on each atom is initially assigned its formal charge. In each iteration, the electronegativities are calculated using Equation (4.39) and hence the charge to be transferred. The total charge on an atom at the end of each iteration is thus obtained by adding the charge transferred from all bonds to the atom to the value of the charge from the previous iteration. The damping factor α^k reduces the influence of the more electronegative atoms. This influence decreases with each iteration. With a damping factor of $\frac{1}{2}$ rapid convergence is achieved, usually within four or five steps.

A somewhat related method is the charge equilibration method of Rappé and Goddard [Rappé and Goddard 1991]. This is employed in the 'Universal Force Field' (UFF) [Rappé *et al.* 1992] as a general method for calculating charge distributions over a very wide range of molecules (in principle, the entire periodic table). An additional feature of the method is that the charges are dependent upon the molecular geometry and so can change during the course of a calculation such as a molecular dynamics simulation. The starting point for this approach is a series expansion of the energy of an isolated atom in terms of the charge:

$$\nu_A(q) = \nu_{A0} + q_A \left(\frac{\partial \nu}{\partial q} \right)_{A0} + \frac{1}{2} q_A^2 \left(\frac{\partial^2 \nu}{\partial q^2} \right)_{A0} + \dots \quad (4.41)$$

Truncating this expansion after second-order terms and considering three specific states (for charges of 0, +1 and -1) leads to:

$$\nu_A(0) = \nu_{A0} \quad (4.42)$$

$$\nu_A(+1) = \nu_{A0} + q_A \left(\frac{\partial \nu}{\partial q} \right)_{A0} + \frac{1}{2} q_A^2 \left(\frac{\partial^2 \nu}{\partial q^2} \right)_{A0} \quad (4.43)$$

$$\nu_A(-1) = \nu_{A0} - q_A \left(\frac{\partial \nu}{\partial q} \right)_{A0} + \frac{1}{2} q_A^2 \left(\frac{\partial^2 \nu}{\partial q^2} \right)_{A0} \quad (4.44)$$

Now the energy of the positive species is the ionisation potential (*IP*) and the energy of the negative species is minus the electron affinity (*EA*). Combining these results gives:

$$\left(\frac{\partial \nu}{\partial q} \right)_{A0} = \frac{1}{2}(IP + EA) = \chi_A^0 \quad (4.45)$$

$$\left(\frac{\partial^2 \nu}{\partial q^2} \right)_{A0} = IP - EA \quad (4.46)$$

As usual, χ_A is the electronegativity. Rappé and Goddard suggested that for a neutral atom with a singly occupied orbital the difference between the ionisation potential and the electron affinity would correspond to the Coulomb repulsion between two electrons placed in that orbital (the orbital would be unoccupied in the positive ion and doubly occupied in the negative species). Writing this difference as J_{AA}^0 (referred to as the *idempotential*) leads to:

$$\nu_A(q) = \nu_{A0} + \chi_A^0 q_A + \frac{1}{2} J_{AA}^0 q_A^2 \quad (4.47)$$

Both the electronegativity and the idempotential can be derived from atomic data, though such atomic data generally need to be corrected for use in molecular systems. In order to use these equations to derive a set of charges for a molecule we first consider the total electrostatic energy of the system:

$$\mathcal{V}(q_1 \cdots q_N) = \sum_{i=1}^N (\nu_{A0} + \chi_A^0 q_A + \frac{1}{2} q_A^2 J_{AA}^0) + \sum_{A=1}^N \sum_{B=A+1}^N q_A q_B J_{AB} \quad (4.48)$$

In this equation J_{AB} represents a formulation of the Coulomb energy between charges q_A and q_B . For well-separated atoms a simple $1/r$ dependency is used. However, this simple Coulomb law is not appropriate for atoms whose charge distributions overlap. In such circumstances (which particularly arise for bonded atoms) there is a significant shielding correction. This shielding correction is a Coulomb integral (Equation (2.107)), with the atomic density being described using a single Slater type orbital whose precise form depends on the nature (ns, np or nd) of the outer valence orbital together with the covalent radius.

In order to derive the actual charges we first incorporate the factors J_{AA}^0 (the limiting value of J_{AA} as the distance tends to zero) into the double summation in Equation (4.48):

$$\mathcal{V}(q_1 \cdots q_N) = \sum_{A=1}^N (\nu_{A0} + \chi_A^0 q_A) + \frac{1}{2} \sum_{A=1}^N \sum_{B=1}^N q_A q_B J_{AB} \quad (4.49)$$

We can then take the derivative of the energy with respect to q_A , which leads to:

$$\frac{\partial \mathcal{V}}{\partial q_A} = \chi_A^0 + \sum_{A=1}^N q_B J_{AB} = \chi_A^0 + J_{AA}^0 q_A + \sum_{B=1, B \neq A}^N q_B J_{AB} \quad (4.50)$$

The derivative of the energy with respect to the charge is an atomic chemical potential; at equilibrium these chemical potentials will all be equal. The electrons move from regions of low electronegativity (high electrochemical potential) to regions of high electronegativity (low electrochemical potential). A further constraint is that the sum of the atomic charges must sum to the total charge on the molecule. These conditions enable a set of simultaneous equations to be written (subject to per-element limits on the charge on any given atom).

The presence of the $q_A q_B$ term with its implied distance dependency means that the charges depend upon the molecular geometry. Thus, should the conformation of a molecule change the atomic charges will also change. Just three parameters are required for each atom in the system (the electronegativity, the idempotential and the covalent radius).

4.9.7 Beyond Partial Atomic Charge Models

Most of the charge models that we have considered so far place the charge on the nuclear centres. Atom-centred charges have many advantages. For example, the electrostatic forces due to charge-charge interactions then act directly on the nuclei. This is important if one wishes to calculate the forces on the nuclei as is required for energy minimisation or a molecular dynamics simulation. Nuclear-centred charges do nevertheless suffer from some drawbacks. In particular, they assume that the charge density about each atom is spherically symmetrical. However, an atom's valence electrons are often distributed in a far from spherical manner, especially in molecules that contain features such as lone pairs and π electron clouds above aromatic ring systems.

4.9.8 Distributed Multipole Models

One way to represent the anisotropy of a molecular charge distribution is to use *distributed multipoles*. In this model, point charges, dipoles, quadrupoles and higher multipoles are distributed throughout the molecule. These distributed multipoles can be determined in various ways but the distributed multipole analysis (DMA) model of A J Stone [Stone 1981; Stone and Alderton 1985] is probably the best-known example. The DMA method calculates the multipoles from a quantum mechanics wavefunction defined in terms of Gaussian basis functions. As we saw in Section 2.6, the overlap between two Gaussian functions can be represented by another Gaussian located at a point (P) along the line that connects them. Each product of basis functions $\phi_\mu \phi_\nu$ thus corresponds to a charge density at P. This density can be expressed as a multipole expansion about P. The highest multipole moment in the local expansion depends upon the basis set used; no multipole moment higher than the sum of the angular quantum numbers of the basis set is possible. Thus, when using a basis set that contains just s and p functions there will be local multipoles no higher than the quadrupole. The crucial feature is that the local multipole expansion

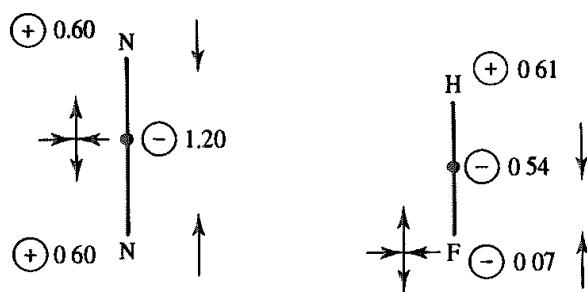


Fig 4.22: Distributed multipole models for N_2 and HF (Figure adapted from Stone A J and M Alderton 1985 *Distributed Multipole Analysis Methods and Applications*. Molecular Physics 56:1047–1064.)

about P can be represented as a multipole expansion about another nearby point S. In the distributed multipole approach, a set of site points is chosen and then the local multipole expansion for each pair of basis functions is ‘moved’ from the relevant point P to one of the sites S.

There are no limitations on the number or location of the multipole sites S; a natural set to use is obtained by placing a site point on each atomic nucleus. In some applications (especially for small molecules) additional sites are defined at the centres of bonds. For example, Stone derived a distributed multipole model for nitrogen from a Dunning [5s4p2d] basis set with two polarisation functions. This model contains charges of +0.60 on the nuclei and a charge of -1.20 at the centre of the bond, together with a dipole on each of the two nuclei and a quadrupole located at the centre of the bond (see Figure 4.22). For HF charges are placed on the two nuclei and at the centre of the bond with a dipole and a quadrupole on the fluorine and a small dipole at the centre of the bond (Figure 4.22). In larger molecules not every atom may be given a site, such as hydrogen atoms bonded to apolar atoms. It is also possible to restrict the order of the multipole expansion at a given atom so that, for example, only a charge component would be present on a polar hydrogen with the higher moments being represented by multipoles on the atom to which it is bonded. An important consideration when choosing the multipole sites is that, when a local multipole expansion is moved, the resulting multipole expansion is no longer a truncated series. However, the smaller the distance between P and the corresponding site point S, the quicker the series converges. In practice, therefore, each local multipole moment expansion is either moved to the nearest site point or is divided between the two nearest site points when they are equally close. With a basis set that contains just s and p functions and multipole sites at the atomic nuclei, it is usually found that the distributed multipole series converges rapidly after the quadrupole term. The multipoles themselves can vary considerably with the basis set used to perform the *ab initio* calculation, but the various electronic properties derived from them usually do not change much.

The distributed multipole model automatically includes non-spherical, anisotropic effects due to features such as lone pairs or π electrons. The original applications of the DMA approach were to small molecules such as diatomics and triatomics. The method has since been used to develop models for nucleic acids and for peptides and has even been applied to the undecapeptide cyclosporin [Price *et al.* 1989], which contains 199 atoms (the

quantum mechanical calculation on this molecule used 1000 basis functions). However, distributed multipole models have not yet been widely incorporated into force fields, not least because of the additional computational effort required. It can be complicated to calculate the atomic forces with the distributed multipole model; in particular, multipoles that are not located on atoms generate torques, which must be analysed further to determine the forces on the nuclei.

4.9.9 Using Charge Schemes to Study Aromatic–Aromatic Interactions

The attractive interactions between molecules containing π systems have long been studied by theoreticians and experimentalists. Such systems are involved in a variety of phenomena, including the stacking of the nucleic acid bases in DNA, the packing of aromatic molecules in crystals and interactions between amino acid side chains in proteins. A variety of orientations are observed for aromatic dimers, ranging from edge-on, T-shaped structures to face-to-face structures (Figure 4.23). Within these two families the molecules can move relative to each other, so that, for example, in a face-to-face arrangement the atoms are overlaid or are staggered. In the T-shaped structure the large quadrupole moments of the benzene molecules adopt their most favourable orientation.

One very simple model of the interactions in such systems was devised by Hunter and Saunders [Hunter and Saunders 1990], who wanted to explain the stacking behaviour of aromatic systems such as the porphyrins shown in Figure 4.24. It is experimentally observed that these molecules adopt a cofacial arrangement with their centres offset as shown. Hunter and Saunders placed point charges not only at the nuclei but also at locations above and below each atom, perpendicular to the plane of the ring. Thus in benzene each carbon atom was given a charge of +1 and also had two associated charges of $-\frac{1}{2}$ above and below the ring (Figure 4.25). The electrostatic interaction between two ring systems is

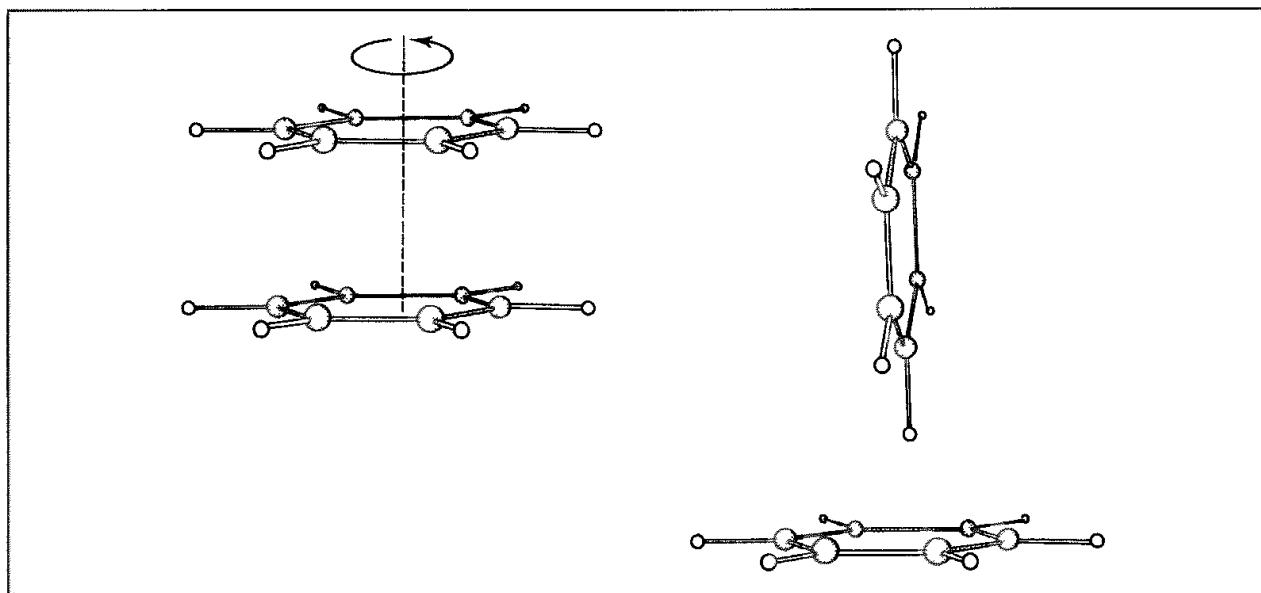


Fig. 4.23 Face-to-face (left) and T-shaped (right) orientations of the benzene dimer

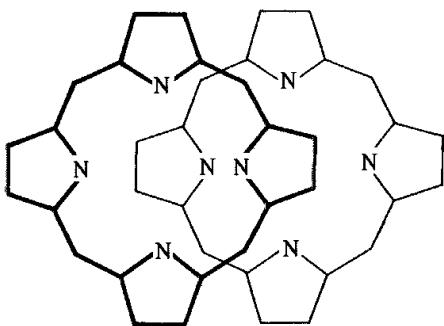


Fig. 4.24: Porphyrin system typical of those studied by Hunter and Saunders [Hunter and Saunders 1990]

calculated in the usual way by summing the charge-charge interactions using Coulomb's law. A major advantage of the Hunter-Saunders approach is its computational simplicity. Moreover, it can be extended to cover a wide range of atom types and so applied to many systems [Vinter 1994] with particular emphasis on simulating DNA [Hunter 1993, Packer *et al.* 2000]. Hunter and Saunders summarised the results of their investigations on porphyrins in three rules:

1. $\pi-\pi$ repulsion dominates in a face-to-face geometry;
2. $\pi-\sigma$ attraction dominates in an edge-on geometry;
3. $\pi-\sigma$ attraction dominates in an offset π -stacked geometry.

The interactions between aromatic systems have also been studied using point charge models, central multipoles and distributed multipoles. Fowler and Buckingham examined homodimers of *sym*-triazine and 1,3,5-trifluorobenzene (Figure 4.26) [Fowler and Buckingham 1991]. They were particularly keen to calculate how the electrostatic energy changed as the rings were twisted in the face-to-face geometry. All but one of the energy models suggested that the staggered orientations were the arrangements of minimum energy, but the energy difference between the eclipsed and staggered structures varied widely, depending upon the model. The central multipole model was found to be ineffective due to convergence problems. Three different point-charge models were considered, all of

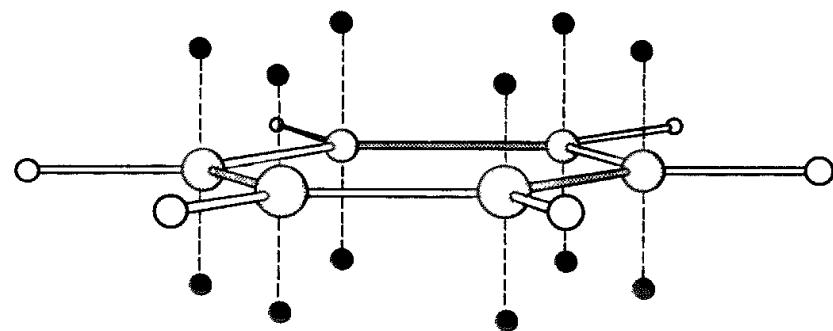


Fig. 4.25: Anisotropic model of benzene developed by Hunter and Saunders [Hunter and Saunders 1990]

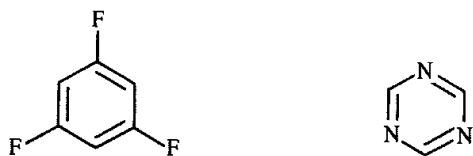


Fig 4.26. Sym-triazine and 1,3,5-trifluorobenzene.

which gave acceptable energy curves. The distributed multipole model also performed well, being comparable to the most accurate of the point-charge models.

4.9.10 Polarisation

Our discussion of electronic effects has concentrated so far on 'permanent' features of the charge distribution. Electrostatic interactions also arise from changes in the charge distribution of a molecule or atom caused by an external field, a process called *polarisation*. The primary effect of the external electric field (which in our case will be caused by neighbouring molecules) is to induce a dipole in the molecule. The magnitude of the induced dipole moment μ_{ind} is proportional to the electric field E , with the constant of proportionality being the polarisability α :

$$\mu_{\text{ind}} = \alpha E \quad (4.51)$$

The energy of interaction between a dipole μ_{ind} and an electric field E (the induction energy) is determined by calculating the work done in charging the field from zero to E , using the following integral:

$$v(\alpha, E) = - \int_0^E dE \mu_{\text{ind}} = - \int_0^E dE \alpha E = - \frac{1}{2} \alpha E^2 \quad (4.52)$$

In strong electric fields contributions to the induced dipole moment that are proportional to E^2 or E^3 can also be important, and higher-order moments such as quadrupoles can also be induced. We will not be concerned with such contributions.

For isolated atoms, the polarisability is isotropic – it does not depend on the orientation of the atom with respect to the applied field, and the induced dipole is in the direction of the electric field, as in Equation (4.51). However, the polarisability of a molecule is often anisotropic. This means that the orientation of the induced dipole is not necessarily in the same direction as the electric field. The polarisability of a molecule is often modelled as a collection of isotropically polarisable atoms. A small molecule may alternatively be modelled as a single isotropic polarisable centre.

Let us consider the electric field due to a dipole μ aligned along the z axis. The magnitude of the electric field at a point P due to the dipole (see Figure 4.27) is:

$$E(r, \theta) = \frac{\mu \sqrt{1 + 3 \cos^2 \theta}}{4\pi\epsilon_0 r^3} \quad (4.53)$$

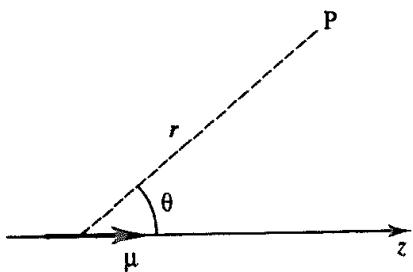


Fig. 4.27: Electric field at point P due to dipole at the origin

The induction energy with another molecule of polarisability α placed at P is therefore

$$\nu(r, \theta) = -\alpha\mu^2 \frac{1 + 3\cos^2\theta}{(4\pi\epsilon_0 r^3)^2} \quad (4.54)$$

The interaction between a dipole and an induced dipole is independent of the disorienting effect of thermal motion, whereas the dipole-dipole interaction between two permanent dipoles does vary with the relative orientation of the two dipoles. This is because the induced dipole follows the direction of the permanent dipole even as the molecules change their orientations as a consequence of molecular collisions.

An important consideration when modelling polarisation effects is that the dipole induced on a molecule (A) will affect the charge distribution of another molecule (B). The electric field at A due to the dipole(s) on B will in turn be affected. The presence of other molecules can also influence the interaction. Consider the polarisation interaction between a polar molecule and a neighbour (Figure 4.28). A third molecule may reduce the size of the electric field on the second molecule and so lower the induction energy. This type of three-body effect will be particularly significant when polarisable atoms are close to polar groups. Polarisation is a cooperative effect and, as such, is modelled using a set of coupled equations which are typically solved iteratively. Initially, the induced dipoles are set to zero. An initial approximation to each induced dipole is then calculated from the permanent charges (i.e. partial atomic charges). The electric field due to these induced dipoles is then added to the electric field from the permanent charges. This gives a refined value of the electric field from which a new induced dipole can be determined. The calculation continues until the induced dipoles do not change significantly between iterations.

A variety of schemes for including polarisation into molecular mechanics force fields have been devised. One approach is to model the polarisation effects at the atomic level, with

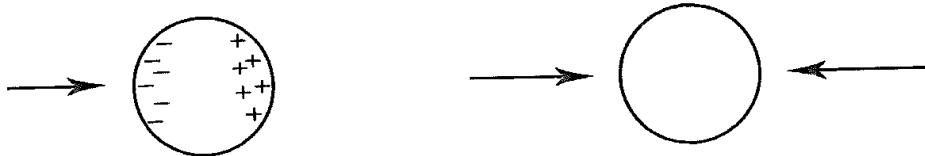


Fig. 4.28: The polarisation interaction between a dipole and a polarisable molecule can be affected by the presence of a second dipole (right) and is therefore a many-body effect

dipoles being induced on each atom [Dang *et al.* 1991]. The magnitude of the dipole induced on an atom i is given by:

$$\mu_{\text{ind},i} = \alpha_i E_i \quad (4.55)$$

α_i is the atomic polarisability, assumed to be isotropic. Appropriate values of α_i have been determined for various systems. The electric field, E_i , at atom i is the vector sum of the field due to the permanent and induced dipoles of the other atoms in the system:

$$E_i = \sum_{j \neq i} \frac{q_j r_{ij}}{r_{ij}^3} + \sum_{j \neq i} \frac{\mu_j}{r_{ij}^3} \left(3r_{ij} \frac{r_{ij}}{r_{ij}^2} - 1 \right) \quad (4.56)$$

r_i and r_j are the position vectors of the atoms i and j . Convergence of these equations in procedures such as molecular dynamics, where successive configurations are generated, can be accelerated if the induced dipoles obtained at each current step are used as the starting points for the next configuration.

An alternative way to model polarisation effects is exemplified by the water model of Sprik and Klein [Sprik and Klein 1988], where the polarisation centre is represented as a collection of closely spaced charges whose values are permitted to vary but whose total sums to zero. In the water model, shown in Figure 4.29, four tetrahedrally arranged charges are used to model the polarisation centre. These charges endow the molecule with an induced dipole moment of any magnitude and direction. The charges are determined iteratively for each configuration of the system. The isotropic polarisability of a simple ion can similarly be treated using two charges of equal magnitude but opposite sign placed either side of the ion. The direction of the 'bond' linking the two polarisation charges and the ion can reorient to change the direction of the induced dipole. In a subsequent refinement of this model Sprik and Klein replaced the point charges by Gaussian charge distributions at the polarisation sites; these were better at modelling features such as hydrogen bonding.

One appealing approach is the dynamically fluctuating charge model of Berne and colleagues [Rick *et al.* 1994]. This method has much in common with the charge equilibration scheme of Rappé and Goddard (see Section 4.9.6) in its use of the electronegativity equalisation approach, which ensures that the atomic chemical potentials are equal in the molecule. The charges are considered as dynamically fluctuating variables, along with the atomic nuclei in a molecular dynamics simulation. This means that the charges evolve in a natural

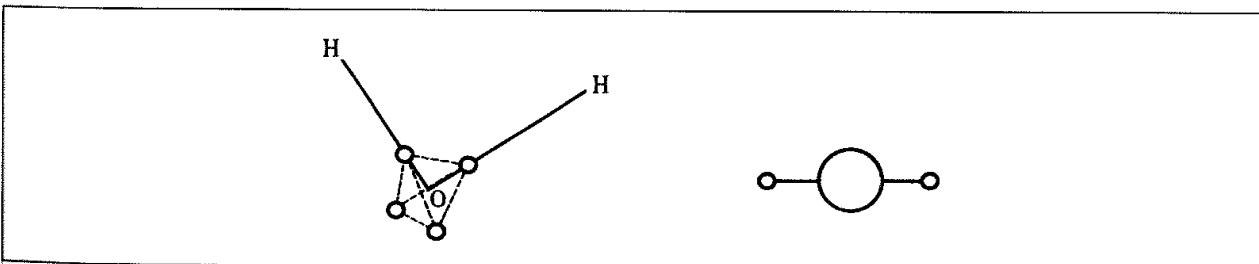


Fig. 4.29. Polarisable models of water and ions developed by Sprik and Klein. (Figure adapted from Sprik M 1993 Effective Pair Potentials and Beyond In Computer Simulation in Chemical Physics, Allen M P, D J Tildesley (Editors) Dordrecht, Kluwer)

manner during the simulation rather than having to determine a new set of charges at each iteration of the procedure. This fluctuating charge model includes intramolecular interactions and so the traditional Coulombic $1/r$ expression is not appropriate. Rather, the charges are replaced by charge distributions (formulated as Slater's orbitals) whose interaction is calculated using a Coulomb integral expression. This interaction is effectively identical to the standard Coulomb expression for intermolecular interactions, only differing for the intramolecular contribution.

One feature of this oscillating charge model is that it requires rather less computational effort than traditional polarisation models. It also implicitly preserves the higher-order multipole terms, which need to be explicitly incorporated in some of the alternative approaches. Ions are represented by two partial charges (which sum to the required integral ionic charge) which are connected by a harmonic spring. The mass of one of these two species is made much greater than the other so that the heavier site remains near the centre of mass as the spring oscillates. This particular model has been used for simulations of pure liquid water [Rick *et al.* 1994], the solvation of amides [Rick and Berne 1996] and to investigate the effects of polarisability on the hydration of the chloride ion in water clusters [Stuart and Berne 1996]. These calculations predicted that the chloride ions were located on the outside of the clusters, even when they contained more than 100 water molecules. This was in contrast to equivalent calculations using a non-polarisable model, the difference being attributed to the presence of fluctuations in the dipole strengths of the water molecules in the cluster, which are, as a consequence, more mobile.

Due to the computational expense, polarisation effects are often included in a calculation only when their effect is likely to be significant, such as simulations of ionic solutions. These systems usually contain atoms or ions and small molecules only. It is important to be aware of the following problem when using atomic polarisabilities. Consider a diatomic molecule. The application of an external field will induce dipoles on both atoms. The dipole on one atom will also contribute to the electric field at the other atom, and thereby influence its induced dipole, but the model takes no account of the fact that the charge distributions on the two atoms are inherently linked. For this reason (and for reasons of computational efficiency) it is common to treat small molecules such as water as single polarisable centres when calculating polarisation effects.

4.9.11 Solvent Dielectric Models

All of the formulae that we have written for electrostatic energies, potentials and forces have included the permittivity of free space, ϵ_0 . This is as one would expect for species acting in a vacuum. However, under some circumstances a different dielectric model is used in the equations for the electrostatic interactions. This is often done when it is desired to mimic solvent effects, without actually including any explicit solvent molecules. One effect of a solvent is to dampen the electrostatic interactions. A very simple way to model this damping effect is to increase the permittivity, most easily by using an appropriate value for the relative permittivity in the Coulomb's law equation (i.e. $\epsilon = \epsilon_0\epsilon_r$). An alternative approach is to make the dielectric dependent upon the separation of the charged species; this gives rise

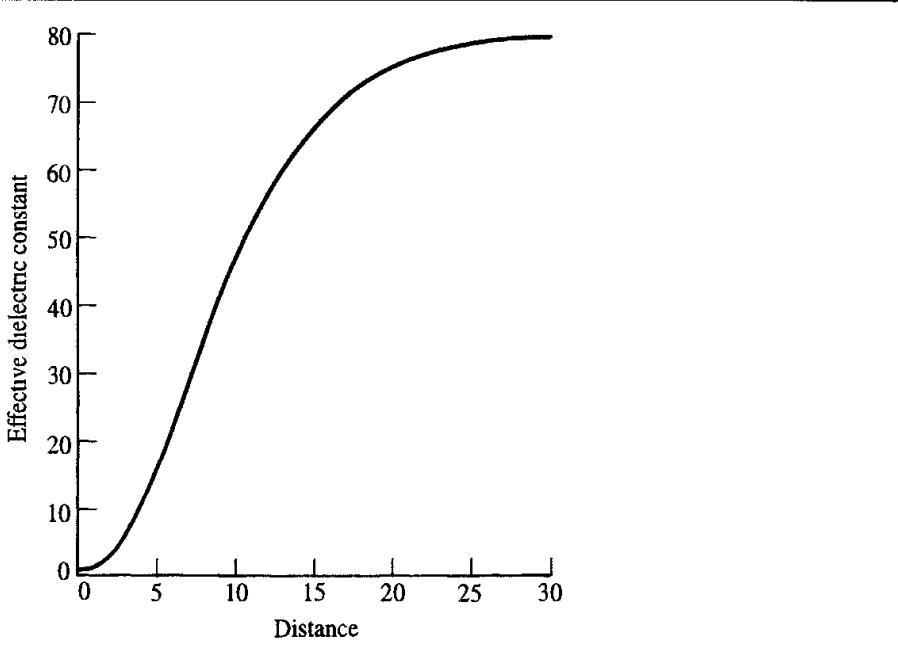


Fig. 4.30 A sigmoidal dielectric model smoothly varies the effective permittivity from 80 to 1 as shown

to the so-called distance-dependent dielectric models. The simplest implementation of a distance-dependent dielectric is to make the relative permittivity proportional to the distance. The interaction energy between two charges q_i and q_j then becomes:

$$\nu(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r^2} \quad (4.57)$$

The simple distance-dependent dielectric has no physical basis and so it is not generally recommended, except when no alternative is possible. More sophisticated distance-dependent functions can also be employed. Many of these have an approximately sigmoidal shape in which the relative permittivity is low at short distances and then rises towards the bulk value at long distances. One example of such a function is [Smith and Pettit 1994]:

$$\epsilon_{\text{eff}}(r) = \epsilon_r - \frac{\epsilon_r - 1}{2} [(rS)^2 + 2rS + 2] e^{-rS} \quad (4.58)$$

The value of ϵ_{eff} varies from a value of 1 at zero separation to ϵ_r (the bulk permittivity of the solvent) at large distances, in a manner determined by the parameter S (which is typically given a value between 0.15 \AA^{-1} and 0.3 \AA^{-1} ; Figure 4.30). Sigmoidal functions give better behaviour than the simple distance-dependent dielectric model. However, it may be difficult to choose the appropriate value for the bulk dielectric ϵ_r when performing calculations on large solutes, as the shortest distance between two charges may be through the solute molecule rather than through the solvent (Figure 4.31).

The polarisation term can be a major contributor to the free energy of solvation of a solute, and a variety of schemes have been devised to incorporate such effects where the solvent is modelled as a continuum. We shall discuss these methods in more detail in Sections 11.9–11.12.

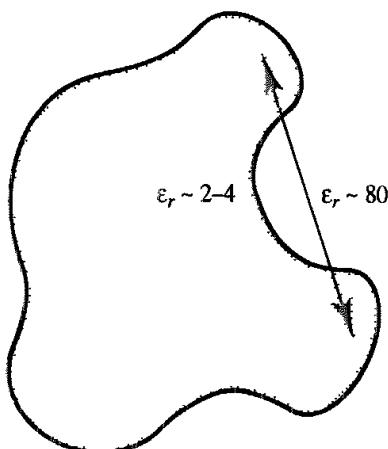


Fig. 4.31 A line joining two points may pass through regions of different permittivity

4.10 Van der Waals Interactions

Electrostatic interactions cannot account for all of the non-bonded interactions in a system. The rare gas atoms are an obvious example; all of the multipole moments of a rare gas atom are zero and so there can be no dipole-dipole or dipole-induced dipole interactions. But there clearly must be interactions between the atoms, how else could rare gases have liquid and solid phases or show deviations from ideal gas behaviour? Deviations from ideal gas behaviour were famously quantitated by van der Waals, thus the forces that give rise to such deviations are often referred to as van der Waals forces.

If we were to study the interaction between two isolated argon atoms using a molecular beam experiment then we would find that the interaction energy varies with the separation in a manner as shown in Figure 4.32. The other rare gases show a similar behaviour. The essential features of this curve are as follows. The interaction energy is zero at infinite distance (and indeed is negligible even at relatively short distances). As the separation is reduced, the energy decreases, passing through a minimum at a distance of approximately 3.8 Å for argon. The energy then rapidly increases as the separation decreases further. The force between the atoms, which equals minus the first derivative of the potential energy with respect to distance, is also shown in Figure 4.32. A variety of experiments have been used to provide evidence for the nature of the van der Waals interactions, including gas imperfections, molecular beams, spectroscopic studies and measurements of transport properties.

4.10.1 Dispersive Interactions

The curve in Figure 4.32 is usually considered to arise from a balance between attractive and repulsive forces. The attractive forces are long-range, whereas the repulsive forces act at short distances. The attractive contribution is due to *dispersive forces*. London first showed how the dispersive force could be explained using quantum mechanics [London 1930] and so this interaction is sometimes referred to as the London force. The dispersive force

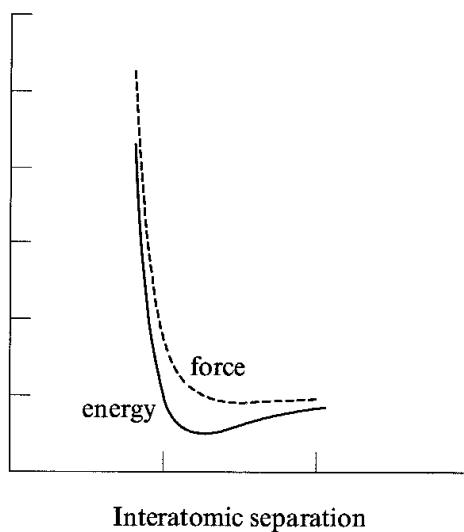


Fig. 4.32: The interaction energy and the force between two argon atoms

is due to instantaneous dipoles which arise during the fluctuations in the electron clouds. An instantaneous dipole in a molecule can in turn induce a dipole in neighbouring atoms, giving rise to an attractive inductive effect.

A simple model to explain the dispersive interaction was proposed by Drude. This model consists of 'molecules' with two charges, $+q$ and $-q$, separated by a distance r . The negative charge performs simple harmonic motion with angular frequency ω along the z axis about the stationary positive charge (Figure 4.33). If the force constant for the oscillator is k and if the mass of the oscillating charge is m , then the potential energy of an isolated Drude molecule is $\frac{1}{2}kz^2$, where z is the separation of the two charges. ω is related to the force constant by $\omega = \sqrt{k/m}$. The Schrödinger equation for a Drude molecule is:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial z^2} + \frac{1}{2}kz^2\psi = E\psi \quad (4.59)$$

This is the Schrödinger equation for a simple harmonic oscillator. The energies of the system are given by $E_\nu = (\nu + \frac{1}{2}) \times \hbar\omega$ and the zero-point energy is $\frac{1}{2}\hbar\omega$.

We now introduce a second Drude molecule, identical to the first, with the positive charge also located on the z axis and an oscillating negative charge (Figure 4.33). When the two molecules are infinitely separated, they do not interact and the total ground-state energy of the system is

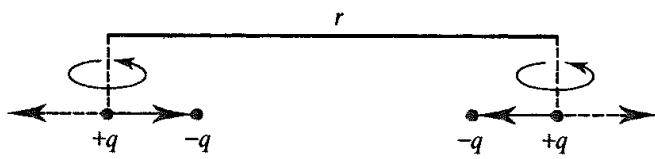


Fig. 4.33 The Drude model for dispersive interactions (Figure adapted from Rigby M, E B Smith, W A Wakeham and G C Maitland 1986 The Forces Between Molecules Oxford, Clarendon Press)

just twice the zero-point energy of a single molecule, $\hbar\omega/2\pi$. As the molecules approach (along the z axis) there are interactions between the two dipoles, and the interaction energy between the two ‘molecules’ can be shown to be approximately given by (see Appendix 4.1):

$$\nu(r) = -\frac{\alpha^4 \hbar\omega}{2(4\pi\epsilon_0)^2 r^6} \quad (4.60)$$

The Drude model thus predicts that the dispersion interaction varies as $1/r^6$.

The two-dimensional Drude model can be extended to three dimensions, the result being:

$$\nu(r) = -\frac{3\alpha^4 \hbar\omega}{4(4\pi\epsilon_0)^2 r^6} \quad (4.61)$$

The Drude model only considers the dipole–dipole interaction; if higher-order terms, due to dipole–quadrupole, quadrupole–quadrupole, etc., interactions are included as well as other terms in the binomial expansion, then the energy of the Drude model is more properly written as a series expansion:

$$\nu(r) = \frac{C_6}{r^6} + \frac{C_8}{r^8} + \frac{C_{10}}{r^{10}} + \dots \quad (4.62)$$

All of the coefficients C_n are negative, implying an attractive interaction. Despite its simplicity, the Drude model gives quite reasonable results; if just the C_6 term is included then for argon the resulting dispersion energy is only about 25% too small.

4.10.2 The Repulsive Contribution

Below about 3 Å, even a small decrease in the separation between a pair of argon atoms causes a large increase in the energy. This increase has a quantum mechanical origin and can be understood in terms of the Pauli principle, which formally prohibits any two electrons in a system from having the same set of quantum numbers. The interaction is due to electrons with the same spin, therefore the short-range repulsive forces are often referred to as *exchange forces*. They are also known as overlap forces. The effect of exchange is to reduce the electrostatic repulsion between pairs of electrons by forbidding them to occupy the same region of space (i.e. the internuclear region). The reduced electron density in the internuclear region leads to repulsion between the incompletely shielded nuclei. At very short internuclear separations, the interaction energy varies as $1/r$ due to this nuclear repulsion, but at larger separations the energy decays exponentially, as $\exp(-2r/a_0)$, where a_0 is the Bohr radius.

4.10.3 Modelling Van der Waals Interactions

The dispersive and exchange-repulsive interactions between atoms and molecules can be calculated using quantum mechanics, though such calculations are far from trivial, requiring electron correlation and large basis sets. For a force field we require a means to model the interatomic potential curve accurately (Figure 4.32), using a simple empirical

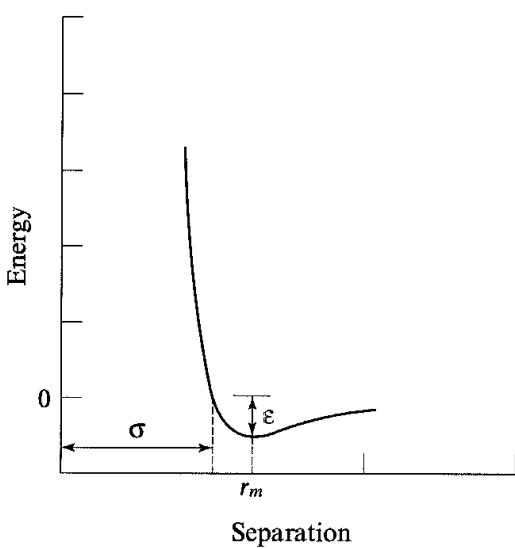


Fig 4.34. The Lennard-Jones potential.

expression that can be rapidly calculated. The need for a function that can be rapidly evaluated is a consequence of the large number of van der Waals interactions that must be determined in many of the systems that we would like to model. The best known of the van der Waals potential functions is the *Lennard-Jones 12-6 function*, which takes the following form for the interaction between two atoms:

$$\nu(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (4.63)$$

The Lennard-Jones 12-6 potential contains just two adjustable parameters: the collision diameter σ (the separation for which the energy is zero) and the well depth ε . These parameters are graphically illustrated in Figure 4.34. The Lennard-Jones equation may also be expressed in terms of the separation at which the energy passes through a minimum, r_m (also written r^*). At this separation, the first derivative of the energy with respect to the internuclear distance is zero (i.e. $\partial\nu/\partial r = 0$), from which it can easily be shown that $r_m = 2^{1/6}\sigma$. We can thus also write the Lennard-Jones 12-6 potential function as follows:

$$\nu(r) = \varepsilon \left\{ \left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right\} \quad (4.64)$$

or

$$\nu(r) = A/r^{12} - C/r^6 \quad (4.65)$$

A is equal to εr_m^{12} (or $4\varepsilon\sigma^{12}$) and C is equal to $2\varepsilon r_m^6$ (or $4\varepsilon\sigma^6$).

The Lennard-Jones potential is characterised by an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12} . These two components are drawn in Figure 4.35. The r^{-6} variation is of course the same power-law relationship found for the leading term in theoretical treatments of the dispersion energy such as the Drude model. There are no strong theoretical arguments in favour of the repulsive r^{-12} , especially as quantum

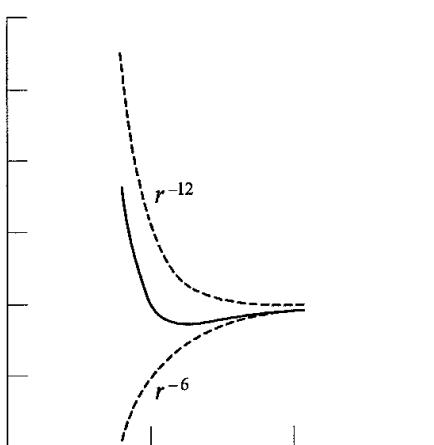


Fig. 4.35 The Lennard-Jones potential is constructed from a repulsive component (αr^{-12}) and an attractive component (αr^{-6})

mechanics calculations suggest an exponential form. The twelfth power term is found to be quite reasonable for rare gases but is rather too steep for other systems such as hydrocarbons. However, the 6–12 potential is widely used, particularly for calculations on large systems, as r^{-12} can be rapidly calculated by squaring the r^{-6} term. The r^{-6} term can also be calculated from the square of the distance without having to perform a computationally expensive square root calculation. Different powers have also been used for the repulsive part of the potential; values of 9 or 10 give a less steep curve and are used in some force fields. Lennard-Jones' original potential has been written in the following general form:

$$\nu(r) = k\epsilon \left[\left(\frac{\sigma}{r} \right)^n - \left(\frac{\sigma}{r} \right)^m \right]; \quad k = \frac{n}{n-m} \left(\frac{n}{m} \right)^{m/(n-m)} \quad (4.66)$$

Equation (4.66) returns the Lennard-Jones potential for $n = 12$ and $m = 6$.

Halgren has proposed an alternative functional form designed to be simple enough to be easily incorporated into molecular mechanics calculations whilst also improving the ability to reproduce experimental data [Halgren 1992, 1996a, b]. In this sense it is an attempt to improve on the Lennard-Jones potential without introducing the complexity of some of the potentials employed by spectroscopists. This potential has the general form:

$$\nu(r) = \epsilon_{ij} \left(\frac{1+\delta}{\rho_{ij} + \delta} \right)^{(n-m)} \left(\frac{1+\gamma}{\rho_{ij}^m + \gamma} - 2 \right) \quad (4.67)$$

In this equation $\rho_{ij} = r_{ij}/r_{ij}^*$. The constants δ and γ apply to all interactions between the atoms i and j . This potential reduces to the standard Lennard-Jones 12–6 potential if the following choice of parameters is used: $n = 12$, $m = 6$, $\delta = \gamma = 0$. Halgren proposed a 'buffered 14–7' potential in which $n = 14$, $m = 7$, $\delta = 0.07$ and $\gamma = 0.12$, giving the following equation:

$$\nu(r) = \epsilon_{ij} \left(\frac{1.07r_{ij}^*}{r_{ij} + 0.07r_{ij}^*} \right)^7 \left(\frac{1.12r_{ij}^7}{r_{ij}^7 + 0.12r_{ij}^{*7}} \right) \quad (4.68)$$

There were several reasons for developing this functional form. First was the desire to keep the potential finite as the interatomic potential approaches zero (unlike the Lennard-Jones function, which becomes infinite). Second, it gives a more accurate reproduction of the series expansion for the dispersion interaction, Equation (4.62). Third, if a larger value of d is used then the repulsive component is greatly reduced without significantly changing the distance at which the potential crosses zero or the depth of the energy minimum. This feature is useful for optimising structures with crude initial geometries; other functional forms can have significant problems with such situations.

In the buffered 14-7 potential the minimum-energy separation r_{ii}^* for an atom i depends on its atomic polarisability:

$$r_{ii}^* = A_i \alpha_i^{1/4} \quad (4.69)$$

Several formulations in which the r^{-12} term in the standard Lennard-Jones formulation is replaced by a theoretically more realistic exponential expression have been proposed. These include the *Buckingham potential*:

$$\nu(r) = \varepsilon \left[\frac{6}{\alpha - 6} \exp[-\alpha(r/r_m - 1)] - \frac{\alpha}{\alpha - 6} \left(\frac{r_m}{r} \right)^6 \right] \quad (4.70)$$

There are three adjustable parameters in the Buckingham potential (ε , r_m and α). A value of α between approximately 14 and 15 gives a potential that closely corresponds to the Lennard-Jones 12-6 potential in the minimum-energy region. When using the Buckingham potential it is important to remember that at very short distances the potential becomes strongly attractive, as shown in Figure 4.36. This could lead to nuclei being fused together during a calculation, and so the program must check that atoms are not becoming too close. The

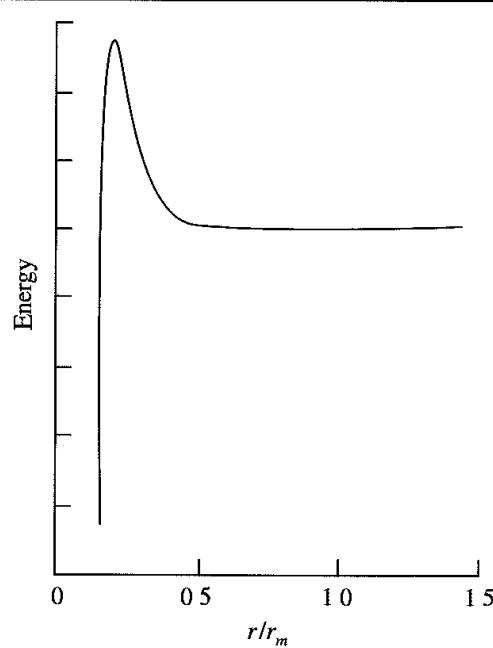


Fig 4.36 A drawback of the Buckingham potential is that it becomes steeply attractive at short distances.

Hill potential is an exponential-6 potential with just two parameters: the minimum energy radius r_m and the well depth ε [Hill 1948]:

$$\nu(r) = -2.25\varepsilon(r_m/r)^6 + 8.28 \times 10^5 \varepsilon \exp(-r/0.0736r_m) \quad (4.71)$$

The Hill potential was originally developed to enable the more realistic exponential term to be written in terms of Lennard-Jones parameters. The coefficients 2.25, 8.25×10^5 and 0.0736 in Equation (4.71) were determined by fitting to data for the rare gases and were assumed to be applicable to other non-polar gases. A Morse potential may also be used to model the van der Waals interactions in a force field, with appropriate parameters.

4.10.4 Van der Waals Interactions in Polyatomic Systems

The interaction energy between molecules depends not only upon their separation but also on their relative orientations and, where appropriate, their conformations. It is usual to calculate the van der Waals interaction energy between two molecules using a site model in which the interaction is determined as the sum of the interactions between all pairs of sites on the two molecules. The sites are often identified with the nuclear positions, but this need not necessarily be the case.

Polyatomic systems invariably involve the calculation of van der Waals interactions between different types of atoms. For example, to calculate the Lennard-Jones interaction energy between two carbon monoxide molecules using a two-site model would require not only van der Waals parameters for the carbon–carbon interactions and the oxygen–oxygen interactions but also for the carbon–oxygen interactions. A system containing N different types of atom would require $N(N - 1)/2$ sets of parameters for the interaction between unlike atoms. The determination of van der Waals parameters can be a difficult and time-consuming process and so it is common to assume that parameters for the cross interactions can be obtained from the parameters of the pure atoms using *mixing rules*. In the commonly used Lorentz–Berthelot mixing rules, the collision diameter σ_{AB} for the A–B interaction equals the arithmetic mean of the values for the two pure species, and the well depth ε_{AB} is given as the geometric mean:

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \quad (4.72)$$

$$\varepsilon_{AB} = \sqrt{\varepsilon_{AA}\varepsilon_{BB}} \quad (4.73)$$

When written in terms of the separation of minimum energy (r^* or r_m), the following notation may be encountered:

$$r_{AB}^* = R_{AA}^* + R_{BB}^* \quad (4.74)$$

R_{AA}^* and R_{BB}^* are atomic parameters, equal to one half of r_{AA}^* and r_{BB}^* , respectively.

The Lorentz–Berthelot combining rules are most successful when applied to similar species. Their major failing is that the well depth can be overestimated by the geometric mean rule. Some force fields calculate the collision diameter for mixed interactions as the geometric mean of the values for the two component atoms. Jorgensen's OPLS force field falls into this category [Jorgensen and Tirado-Reeves 1988].

For the buffered 14–7 functional form more elaborate combination rules are employed:

$$r_{ij}^* = \frac{(r_{ii}^{*3} + r_{jj}^{*3})}{(r_{ii}^{*2} + r_{jj}^{*2})} \quad (4.75)$$

This is similar in spirit to the arithmetic-mean rule but with each individual r_{ii}^* being weighted according to the square of its value. The well depth in this function starts with a formula proposed by Slater and Kirkwood for the C_6 coefficient of the dispersion series expansion:

$$C_{6ij} = \frac{3}{2} \frac{\alpha_i \alpha_j}{(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}} = \frac{2\alpha_i \alpha_j}{\alpha_i^2 C_{ijj} + \alpha_j^2 C_{6ii}} \quad (4.76)$$

In this equation N represents the effective number of electrons and α are atomic polarisabilities; the second formulation in Equation (4.76) is derived using the relationship:

$$N_i = 16C_{6ii}^2/9\alpha_i^3 \quad (4.77)$$

From this the well depths ε are then obtained as follows:

$$\varepsilon_{ij} = \frac{1}{2} \frac{kG_i G_j C_{6ij}}{r_{ij}^{*6}} = \frac{181.16 G_i G_j \alpha_i \alpha_j}{(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}} \frac{1}{r_{ij}^{*6}} \quad (4.78)$$

Here, k is a factor which converts to units (kcal/mol in this case where the distances are in Å and the polarisabilities in Å³). G_i and G_j are constants chosen to reproduce the well depths for like-with-like interactions. The atomic polarisability values are obtained from an examination of appropriate molecular experimental data (such as measurements of molar refractivity).

In some force fields the interaction sites are not all situated on the atomic nuclei. For example, in the MM2, MM3 and MM4 programs, the van der Waals centres of hydrogen atoms bonded to carbon are placed not at the nuclei but are approximately 10% along the bond towards the attached atom. The rationale for this is that the electron distribution about small atoms such as oxygen, fluorine and particularly hydrogen is distinctly non-spherical. The single electron from the hydrogen is involved in the bond to the adjacent atom and there are no other electrons that can contribute to the van der Waals interactions. Some force fields also require lone pairs to be defined on particular atoms; these have their own van der Waals and electrostatic parameters.

The van der Waals and electrostatic interactions between atoms separated by three bonds (i.e. the 1,4 atoms) are often treated differently from other non-bonded interactions. The interaction between such atoms contributes to the rotational barrier about the central bond, in conjunction with the torsional potential. These 1,4 non-bonded interactions are often scaled down by an empirical factor; for example, a factor of 2.0 is suggested for both the electrostatic and van der Waals terms in the 1984 AMBER force field (a scale factor of 1/1.2 is used for the electrostatic terms in the 1995 AMBER force field). There are several reasons why one would wish to scale the 1,4 interactions. The error associated with the use of an r^{-12} repulsion term (which is too steep compared with the more correct exponential term) would be most significant for 1,4 atoms. In addition, when two 1,4

atoms come close together some redistribution of the charge along the connecting bonds would be expected that would act to reduce the interaction. Such a charge redistribution would not be possible for two atoms at a similar distance apart if they were in different molecules.

The parameters for the van der Waals interactions can be obtained in a variety of ways. In the early force fields, such parameters were often determined from an analysis of crystal packing. The objective of such studies was to produce a set of van der Waals parameters which enabled the experimental geometries and thermodynamic properties such as the heat of sublimation to be reproduced as accurately as possible. More recent force fields derive their van der Waals parameters using liquid simulations in which the parameters are optimised to reproduce a range of thermodynamic properties such as the densities and enthalpies of vaporisation for appropriate liquids.

4.10.5 Reduced Units

The Lennard-Jones potential is completely specified by the two parameters ε and σ . This means that the results of a calculation performed on (say) liquid argon can be easily converted to give equivalent results for another noble gas. For this reason it is common to simulate the rare gases in terms of reduced units with ε and σ both set to 1. The results can then be converted to any system as appropriate. For example, the reduced density ρ^* is related to the real density by $\rho^* = \rho\sigma^3$; the reduced energy E^* is given by $E^* = E/\varepsilon$, and so on. Electrostatic interactions given by Coulomb's law are also often written in terms of a reduced unit of charge, which corresponds to each charge being divided by $\sqrt{4\pi\varepsilon_0}$. This means that Coulomb's law takes the less cumbersome form:

$$\nu(q_1, q_2) = q_1 q_2 / r_{12} \quad \text{or} \quad \nu(q_1, q_2) = q_1 q_2 / \varepsilon_r r_{12} \quad (4.79)$$

4.11 Many-body Effects in Empirical Potentials

The electrostatic and van der Waals energies that we have considered so far are calculated between pairs of interaction sites. The total non-bonded interaction energy is thus determined by adding together the interactions between all pairs of sites in the system. However, the interaction between two molecules can be affected by the presence of a third, fourth or more molecules. For example, the interaction energy between three molecules A, B and C is not in general given by the sum of the pairwise interaction energies: $\nu(A, B, C) \neq \nu(A, B) + \nu(A, C) + \nu(B, C)$. We have already seen an example of a non-pairwise contribution, namely the polarisation interaction, which is determined using a self-consistent procedure.

Three-body effects can significantly affect the dispersion interaction. For example, it is believed that three-body interactions account for approximately 10% of the lattice energy of crystalline argon. For very precise work, interactions involving more than three atoms may have to be taken into account, but they are usually small enough to be ignored. A potential that includes both two- and three-body interactions would be written in the following

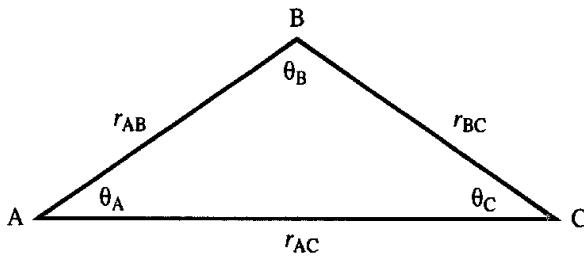


Fig. 4.37: Calculating the three-body Axilrod-Teller contribution

general form:

$$\mathcal{V}(\mathbf{r}^N) = \sum_{i=1}^N \sum_{j=i+1}^N \nu^{(2)}(r_{ij}) + \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N \nu^{(3)}(r_{ij}, r_{ik}, r_{jk}) \quad (4.80)$$

Axilrod and Teller investigated the three-body dispersion contribution and showed that the leading term is:

$$\nu^{(3)}(r_{AB}, r_{AC}, r_{BC}) = \nu_{A,B,C} \frac{3 \cos \theta_A \cos \theta_B \cos \theta_C}{(r_{AB} r_{AC} r_{BC})^3} \quad (4.81)$$

θ_A , θ_B and θ_C are the internal angles of the triangle with sides of length r_{AB} , r_{AC} and r_{BC} (Figure 4.37). $\nu_{A,B,C}$ is a constant characteristic of the three species A, B and C. If A, B and C are identical then $\nu_{A,B,C}$ is approximately related to the Lennard-Jones coefficient C_6 and the polarisability by

$$\nu_{A,B,C} = -\frac{3\alpha C_6}{4(4\pi\epsilon_0)} \quad (4.82)$$

The effect of the Axilrod-Teller term (also known as the triple-dipole correction) is to make the interaction energy more negative when three molecules are linear but to weaken it when the molecules form an equilateral triangle. This is because the linear arrangement enhances the correlations of the motions of the electrons, whereas the equilateral arrangement reduces it.

The three-body contribution may also be modelled using a term of the form $\nu^{(3)}(r_{AB}, r_{AC}, r_{BC}) = K_{A,B,C} \{ \exp(-\alpha r_{AB}) \exp(-\beta r_{AC}) \exp(-\gamma r_{BC}) \}$ where K , α , β and γ are constants describing the interaction between the atoms A, B and C. Such a functional form has been used in simulations of ion-water systems, where polarisation alone does not exactly model configurations when there are two water molecules close to an ion [Lybrand and Kollman 1985]. The three-body exchange repulsion term is thus only calculated for ion-water-water trimers when the species are close together.

The computational effort is significantly increased if three-body terms are included in the model. Even with a simple pairwise model, the non-bonded interactions usually require by far the greatest amount of computational effort. The number of bond, angle and torsional terms increases approximately with the number of atoms (N) in the system, but the number of non-bonded interactions increases with N^2 . There are $N(N - 1)/2$ distinct pairs of

interactions to evaluate for a pairwise potential. If three-body effects are included then there are $N(N - 1)(N - 2)/6$ unique three-body interactions. A system with 1000 atoms has 499 500 pairwise interactions and 166 167 000 three-body interactions. In general, there are approximately $N/3$ times more three-body terms than two-body terms and so it is clear why it is often considered preferable to avoid calculating the three-body interactions.

4.12 Effective Pair Potentials

Fortunately, it is found that a significant proportion of the many-body effects can be incorporated into a pairwise model, if properly parametrised. The pair potentials most commonly used in molecular modelling are thus ‘effective’ pairwise potentials; they do not represent the true interaction energy between two isolated particles but are parametrised to include many-body effects in the pairwise energy. Similarly, polarisation effects can be implicitly included in a force field by the simple expedient of enhancing the electrostatic interaction. This can be done by using larger partial charges than those for an isolated molecule. This is most obviously manifested in larger multipole moments; the dipole moment of a single water molecule is 1.85 D, whereas the dipole moment of many simple water models designed to simulate liquid water are significantly larger (closer to the experimental value for liquid water of 2.6 D).

A notable example of a potential that does include many-body terms is the Barker–Fisher–Watts potential for argon, which combines a pairwise potential with an Axilrod–Teller triple

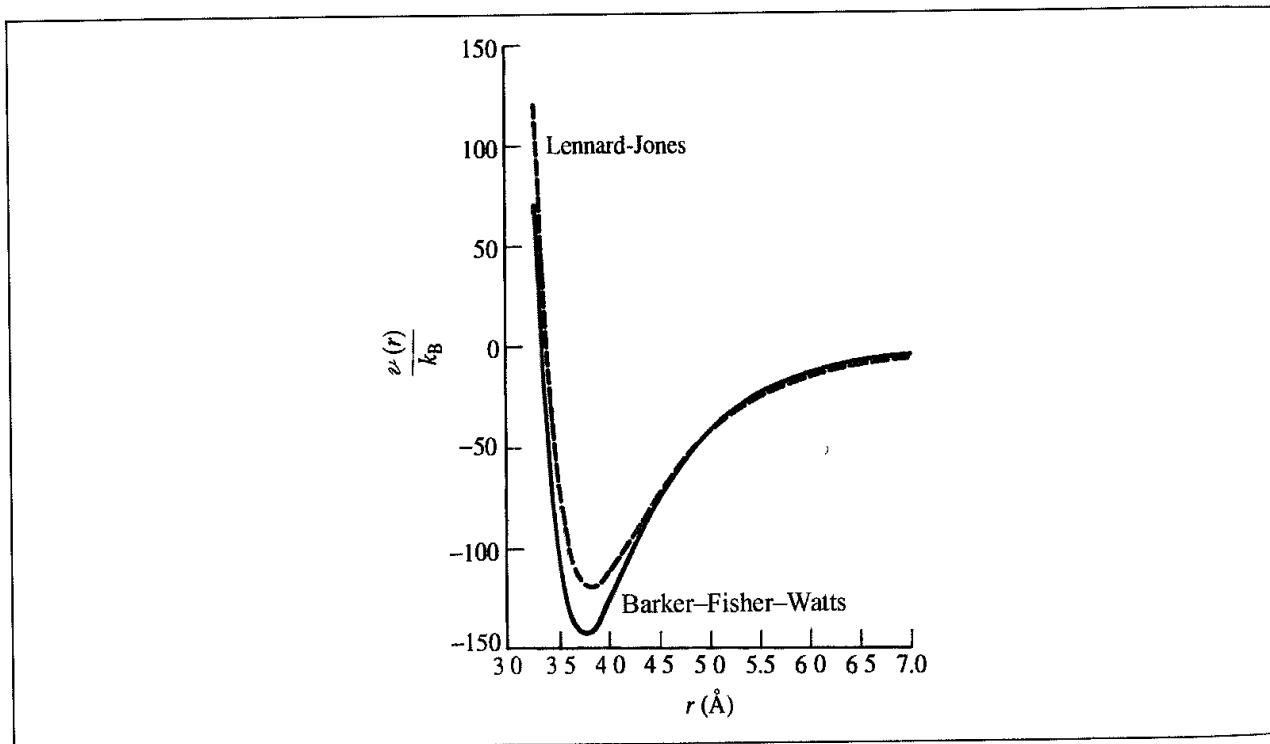


Fig. 4.38. Comparison of the Lennard-Jones potential for argon with the Barker–Fisher–Watts pair potential; k_B is Boltzmann’s constant

potential [Barker *et al.* 1971]. The pair potential is a linear combination of two potentials that each take the following form:

$$\begin{aligned} \nu^*(r) = & e^{\alpha(1-r^*)}[A_0 + A_1(r^* - 1) + A_2(r^* - 1)^2 + A_3(r^* - 1)^3 + A_4(r^* - 1)^4 + A_5(r^* - 1)^5] \\ & + \frac{C_6}{\delta + R^{*6}} + \frac{C_8}{\delta + R^{*8}} + \frac{C_{10}}{\delta + R^{*10}} \end{aligned} \quad (4.83)$$

This potential function contains eleven constants: α , $A_0 \dots A_5$, C_6 , C_8 , C_{10} and δ . The function is expressed in terms of r^* , which is given by $r^* = r/r_m$, where r_m is the separation at the minimum in the potential. The 'true' interaction energy as a function of the separation, r , is then obtained by multiplying $\nu^*(r^*)$ by the depth of the potential well, ε :

$$\nu(r) = \varepsilon \nu^*(r^*) \quad (4.84)$$

A comparison of the pairwise contribution to the Barker–Fisher–Watts potential with the Lennard–Jones potential for argon is shown in Figure 4.38.

4.13 Hydrogen Bonding in Molecular Mechanics

Some force fields replace the Lennard–Jones 6–12 term between hydrogen-bonding atoms by an explicit hydrogen-bonding term, which is often described using a 10–12 Lennard–Jones potential:

$$\nu(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}} \quad (4.85)$$

This function is used to model the interaction between the donor hydrogen atom and the heteroatom acceptor atom. Its use is intended to improve the accuracy with which the geometry of hydrogen-bonding systems is predicted. Other force fields incorporate a more complicated hydrogen-bonding function that takes into account deviations from the geometry of the hydrogen bond and is thus dependent upon the coordinates of the donor and acceptor atoms as well as the hydrogen atom. For example, the YETI force field [Vedani 1988] uses the following form for its hydrogen bonding term:

$$\nu_{\text{HB}} = \left(\frac{A}{r_{\text{H Acc}}^{12}} - \frac{C}{r_{\text{H Acc}}^{10}} \right) \cos^2 \theta_{\text{Don H Acc}} \cos^4 \omega_{\text{H Acc-LP}} \quad (4.86)$$

The energy in Equation (4.86) depends upon the distance from the hydrogen to the acceptor, the angle subtended at the hydrogen by the bonds to the donor and the acceptor, and the deviation of the hydrogen bond from the closest lone-pair direction at the acceptor atom ($\omega_{\text{H Acc-LP}}$ in Equation (4.86), Figure 4.39).

The GRID program [Goodford 1985] that is used for finding energetically favourable regions in protein binding sites uses a direction-dependent 6–4 function:

$$\nu_{\text{HB}} = \left(\frac{C}{d^6} - \frac{D}{d^4} \right) \cos^m \theta \quad (4.87)$$

θ is the angle subtended at the hydrogen and m is usually set to 4.

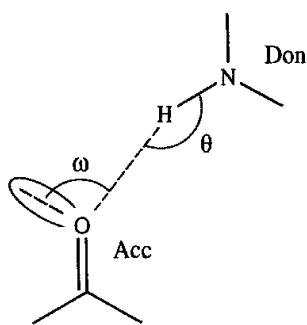


Fig 4.39: Definition of hydrogen-bond geometry used in YETI force field

By no means do all force fields contain explicit hydrogen-bonding terms; most rely upon electrostatic and van der Waals interactions to reproduce hydrogen bonding.

4.14 Force Field Models for the Simulation of Liquid Water

Many of the concepts that we have considered so far can be illustrated by examining some of the empirical models that have been developed to study water. Despite its small size, water acts as a paradigm for the different force field models that we have discussed. Moreover, many of its properties can be easily determined using computer simulation methods and so readily compared with experiment. It is also one of the most challenging systems to model accurately. A wide range of water models have been proposed. The computational efficiency with which the energy can be calculated using a given model is often an important factor as there may be a very large number of water molecules present, together with a solute; most of the force fields used to simulate liquid water thus use effective pairwise potentials with no explicit three-body terms or polarisation effects.

Water models can be conveniently divided into three types. In the simple interaction-site models each water molecule is maintained in a rigid geometry and the interaction between molecules is described using pairwise Coulombic and Lennard-Jones expressions. Flexible models permit internal changes in conformation of the molecule. Finally, models have been developed that explicitly include the effects of polarisation and many-body effects.

4.14.1 Simple Water Models

The ‘simple’ water models use between three and five interaction sites and a rigid water geometry. The TIP3P [Jorgensen *et al.* 1983] and SPC [Berendsen *et al.* 1981] models use a total of three sites for the electrostatic interactions; the partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge located on the oxygen atom. The van der Waals interaction between two water molecules is computed using a Lennard-Jones function with just a single interaction point per molecule centred on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated. The TIP3P and SPC models differ slightly in the geometry of each water molecule, in the

	SPC	SPC/E	TIP3P	BF	TIP4P	ST2
$r(\text{OH})$, Å	1.0	1.0	0.9572	0.96	0.9572	1.0
HOH, deg	109.47	109.47	104.52	105.7	104.52	109.47
$A \times 10^{-3}$, kcal Å ¹² /mol	629.4	629.4	582.0	560.4	600.0	238.7
C , kcal Å ⁶ /mol	625.5	625.5	595.0	837.0	610.0	268.9
$q(\text{O})$	-0.82	-0.8472	-0.834	0.0	0.0	0.0
$q(\text{H})$	0.41	0.4238	0.417	0.49	0.52	0.2375
$q(\text{M})$	0.0	0.0	0.0	-0.98	-1.04	-0.2375
$r(\text{OM})$, Å	0.0	0.0	0.0	0.15	0.15	0.8

Table 4.3 A comparison of various water models [Jorgensen *et al.* 1983]. For the ST2 potential, $q(\text{M})$ is the charge on the 'lone pairs', which are a distance 0.8 Å from the oxygen atom (see Figure 4.40)

hydrogen charges and in the Lennard-Jones parameters. These differences are indicated in Table 4.3, which also includes data for the SPC/E model [Berendsen *et al.* 1987], which is an updated version of the SPC model. The four-site models such as that of Bernal and Fowler [Bernal and Fowler 1933] (which is now relatively little used but is important for historical reasons as it dates from 1933) and Jorgensen's TIP4P model [Jorgensen *et al.* 1983] shift the negative charge from the oxygen atom to a point along the bisector of the HOH angle towards the hydrogens (Figure 4.40). The parameters for these two models are also given in the table. The most commonly used five-site model is the ST2 potential of Stillinger and Rahman [Stillinger and Rahman 1974]. Here, charges are placed on the hydrogen atoms and on two lone-pair sites on the oxygen. The electrostatic contribution is modulated so that for oxygen-oxygen distances below 2.016 Å it is zero and for distances greater than 3.1287 Å it takes its full value. Between these two distances the electrostatic contribution is modulated using a function that smoothly varies from 0.0 at the shorter distance to 1.0 at the longer distance (see Section 6.7.3).

The experimentally determined dipole moment of a water molecule in the gas phase is 1.85 D. The dipole moment of an individual water molecule calculated with any of these simple models is significantly higher; for example, the SPC dipole moment is 2.27 D and that for TIP4P is 2.18 D. These values are much closer to the effective dipole moment of liquid water, which is approximately 2.6 D. These models are thus all effective pairwise models. The simple water models are usually parametrised by calculating various properties using molecular dynamics or Monte Carlo simulations and then modifying the

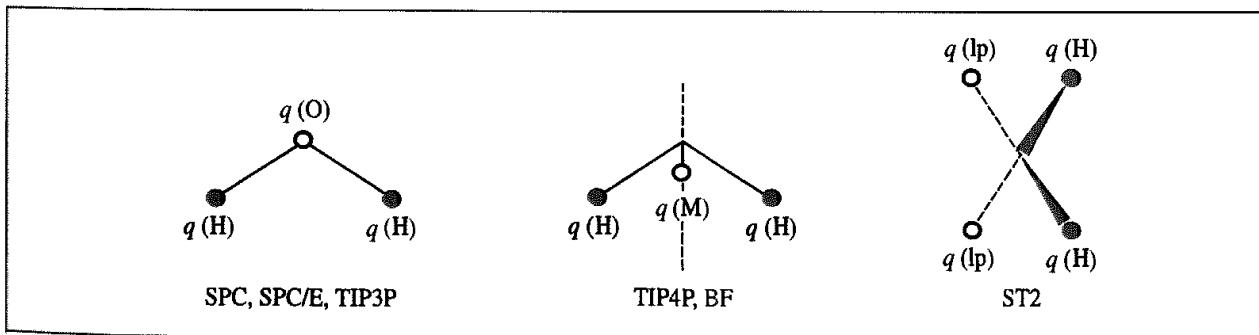


Fig 4.40. Some 'simple' water models (Table 4.3) [Jorgensen *et al.* 1983].

parameters until the desired level of agreement between experiment and theory is achieved. Thermodynamic and structural properties are usually used in the parametrisation, such as the density, radial distribution function, enthalpy of vaporisation, heat capacity, diffusion coefficient and dielectric constant.* It is found that some properties such as the density and the enthalpy of vaporisation are predicted rather well by all of the models, but there is significant variation in the values for other properties such as the dielectric constant [Jorgensen *et al.* 1983]. When comparing the different models, it is also important to take account of the computational effort each requires. Thus, nine site-site distances must be calculated for each water dimer using a three-site model; ten are required for a four-site model, and seventeen for the ST2 model.

The use of a rigid model for water is obviously an approximation, and it means that some properties cannot be determined at all. For example, only when internal flexibility is included can the vibrational spectrum be calculated and compared with experiment. Flexibility is most easily incorporated by 'grafting' bond-stretching and angle-bending terms onto the potential function for a rigid model. Such an approach needs to be done with care. For example, Ferguson has developed a flexible model for water that is based upon the SPC model [Ferguson 1995]. The partial charges and van der Waals parameters in this model were slightly different from those in the rigid model, and flexibility was achieved using cubic and harmonic bond-stretching terms and a harmonic angle-bending term. The calculated values compared well with experimental results for a wide range of thermodynamic and structural properties, including the dielectric constant and self-diffusion coefficient.

4.14.2 Polarisable Water Models

The simple models give very good results for a wide range of properties of pure liquid water. However, there is some concern that they are not appropriate models to use for the most accurate work. This is especially the case for inhomogeneous systems where there are strong electric field gradients due to the presence of ions, and at the solute-solvent interface. Under such circumstances models that explicitly include polarisation effects and three-body terms are considered to be more appropriate. The inclusion of an explicit polarisation term should also enhance the ability of the model to reproduce the behaviour of water in other phases (e.g. solid and vapour) and at the interface between different phases. The dipole moment of an isolated water molecule in such a model should thus be closer to the gas-phase value rather than to the 'effective' value in liquid water. The simplest way to include polarisation is to use an isotropic molecular polarisability contribution, an alternative is to use atom-centred polarisabilities or the variable charge method. The incorporation of polarisability may significantly increase the computational effort required for a liquid simulation, and even then only the best polarisable models currently compete with the well-established models that use effective pairwise potentials. We have already considered some of the polarisable water models in our discussion of polarisation effects. One early attempt to incorporate such effects into a water model was made by Barnes,

* A discussion of the calculation of these properties from computer simulation is given in Section 6.2

Finney, Nicholas and Quinn [Barnes *et al.* 1979]. Their polarisable electropole water model represented the charge distribution by a multipole expansion comprising a dipole of 1 855 D and a quadrupole moment that was determined from quantum mechanical calculations on an isolated molecule. Polarisation effects were calculated using an isotropic molecular polarisability from the electric fields being produced by the dipoles and quadrupoles of surrounding molecules. The model also used a spherically symmetric Lennard-Jones function. A more recent study used the fluctuating charge model with both the TIP4P and SPC geometries [Rick *et al.* 1994]. The charges were assigned to reproduce the correct dipole moment of the gas-phase molecule (in contrast to the equivalent non-polarisable models). Of the two geometries, the TIP4P model gave the better results for various properties. The dielectric properties were considered particularly well reproduced, including features in the dielectric spectrum arising from the translational motion of a water molecule in the cage of its neighbours. This feature is not present in fixed-charge models. Moreover, the computational cost with this particular model was only about 1.1 times that of the fixed-charge equivalent.

4.14.3 *Ab initio* Potentials for Water

The final category of water model that we shall consider are the '*ab initio*' potentials. These are based upon *ab initio* quantum mechanical calculations on small clusters of water molecules. One example of this type is the NCC model of Nieser, Corongiu and Clementi, which combines a two-molecule potential with a polarisation term [Niesar *et al.* 1990]. They had previously tried to explicitly include both three- and four-body effects but found this model computationally too expensive. The two-body model uses partial charges on the hydrogen atoms and a compensating negative charge on a site located along the bisector of the HOH angle, as in the TIP4P model. The equation used is:

$$\begin{aligned} \mathcal{V}_{\text{two-body}} = & q^2 \left(\frac{1}{R_{13}} + \frac{1}{R_{14}} + \frac{1}{R_{23}} + \frac{1}{R_{24}} \right) \\ & + \frac{4q^2}{R_{78}} - 2q^2 \left(\frac{1}{R_{81}} + \frac{1}{R_{82}} + \frac{1}{R_{73}} + \frac{1}{R_{74}} \right) \\ & + A_{OO} e^{-B_{OO}R_{56}} + A_{HH} (e^{-B_{HH}R_{13}} + e^{-B_{HH}R_{14}} + e^{-B_{HH}R_{23}} + e^{-B_{HH}R_{24}}) \\ & + A_{OH} (e^{-B_{OH}R_{53}} + e^{-B_{OH}R_{54}} + e^{-B_{OH}R_{61}} + e^{-B_{OH}R_{62}}) \\ & - A'_{OH} (e^{-B'_{OH}R_{53}} + e^{-B'_{OH}R_{54}} + e^{-B'_{OH}R_{61}} + e^{-B'_{OH}R_{62}}) \\ & + A_{PH} (e^{-B_{PH}R_{73}} + e^{-B_{PH}R_{74}} + e^{-B_{PH}R_{81}} + e^{-B_{PH}R_{82}}) \\ & + A_{PO} (e^{-B_{PO}R_{76}} + e^{-B_{PO}R_{85}}) \end{aligned} \quad (4.88)$$

The points P are the locations where the negative charge is placed (numbered 7 and 8 in Figure 4.41) and the terms A_{PH} and A_{PO} are used to enhance the performance of the model at short distances. q is the charge on each hydrogen. The polarisation term is calculated in an iterative manner using induced dipoles along each O–H bond. The NCC model was parametrised by fitting to the energies and other properties of 250 trimer and

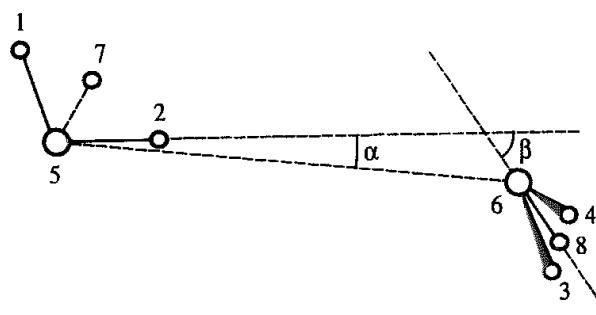


Fig 4.41 The NCC water model (After Corongiu G 1992 Molecular Dynamics Simulation for Liquid Water Using a Polarisable and Flexible Potential International Journal of Quantum Chemistry 42:1209–1235)

350 dimer configurations determined with high-level *ab initio* methods and large basis sets. The water trimer data was used to fit the many-body parameters (i.e. the locations of the induced dipole moments and the point charges, together with the polarisability and the value of the hydrogen charge). The dimer data were then used to fit the remaining terms in the potential.

The original NCC potential was designed as a rigid water model and performed well in tests of its ability to reproduce experimental data for both water dimers and liquid water. A flexible version has also been developed [Corongiu 1992], with the energy being expressed as a function of the three internal coordinates (two bond lengths and one angle) with terms up to quartics:

$$\begin{aligned}
 V_{\text{intra}} = & \frac{1}{2} f_{RR}(\delta_1^2 + \delta_2^2) + \frac{1}{2} f_{\theta\theta}(\delta_3^2) + f_{RR'}\delta_1\delta_2 + f_{R\theta}(\delta_1 + \delta_2)\delta_3 \\
 & + \frac{1}{R_e} [f_{RRR}(\delta_1^3 + \delta_2^3) + f_{\theta\theta\theta}\delta_3^3 + f_{RRR'}(\delta_1 + \delta_2)\delta_1\delta_2 \\
 & \quad + f_{RR\theta}(\delta_1^2 + \delta_2^2)\delta_3 + f_{RR'\theta}\delta_1\delta_2\delta_3 + f_{R\theta\theta}(\delta_1 + \delta_2)\delta_3^2] \\
 & + \frac{1}{R_e^2} [f_{RRRR}(\delta_1^4 + \delta_2^4) + f_{\theta\theta\theta\theta}\delta_3^4 + f_{RRR'R'}(\delta_1^2 + \delta_2^2)\delta_1\delta_2 \\
 & \quad + f_{RRR'R'}\delta_1^2\delta_2^2 + f_{RRR\theta}(\delta_1^3 + \delta_2^3)\delta_3] \\
 & + \frac{1}{R_e^2} [f_{RRR'\theta}(\delta_1 + \delta_2)\delta_1\delta_2\delta_3 + f_{RR\theta\theta}(\delta_1^2 + \delta_2^2)\delta_3^2 \\
 & \quad + f_{RR'\theta\theta}\delta_1\delta_2\delta_3 + f_{R\theta\theta\theta}(\delta_1 + \delta_2)\delta_3^2]
 \end{aligned} \tag{4.89}$$

where $\delta_1 = R_1 - R_e$, $\delta_2 = R_2 - R_e$ and $\delta_3 = R_e(\theta - \theta_e)$.

The functional form of the NCC model demonstrates the complexity of some empirical models (and this for a molecule that contains only three atoms!). We should also note that the development of empirical models from *ab initio* quantum mechanical data is an approach that is already well established and looks likely to be a method that is more widely used in the future.

4.15 United Atom Force Fields and Reduced Representations

In our discussion so far, we have assumed that all of the atoms in the system are explicitly represented in the model. However, as the number of non-bonded interactions scales with the square of the number of interaction sites present, there are clear advantages if the number of interaction sites can be reduced. The simplest way to do this is to subsume some or all of the atoms (usually just the hydrogen atoms) into the atoms to which they are bonded. A methyl group would then be modelled as a single 'pseudo-atom' or 'united atom'. The van der Waals and electrostatic parameters would be modified to take account of the adjoining hydrogen atoms. Considerable computational savings are possible; for example, if butane is modelled as a four-site model rather than one with twelve atoms then the van der Waals interaction between two butane molecules involves the calculation of sixteen terms rather than 144. Other hydrocarbons are often represented using united atom models. Many of the earliest calculations on proteins used united atom representations. In this case, not all of the hydrogen atoms in the protein are subsumed into their adjacent atoms, but just those that are bonded to carbon atoms. Hydrogen atoms bonded to polar atoms such as nitrogen and oxygen are able to participate in hydrogen-bonding interactions, which are modelled much better if these hydrogens are explicitly represented.

One drawback with a united atom force field is that chiral centres may be able to invert during a calculation. This was found to be a problem with the united atom force fields for proteins. The alpha carbon in the peptide unit (C_α in Figure 4.42) is bonded to a hydrogen atom and to the side chain (glycine and proline are slightly different; see Section 10.1). A united atom force field model would not explicitly include the alpha hydrogen. Unfortunately, the stereochemistry at the alpha carbon can then invert during a calculation. This should be avoided as the naturally occurring amino acids have a defined stereochemistry (as shown in Figure 4.42). This inversion may be prevented through the use of an improper torsion term (e.g. $N-C-C_\alpha-R$) to keep the side chain in the correct relative position.

In a united atom force field the van der Waals centre of the united atom is usually associated with the position of the heavy (i.e. non-hydrogen) atom. Thus, for a united CH_3 or CH_2 group the van der Waals centre would be located at the carbon atom. It would be more accurate to associate the van der Waals centre with a position that was offset slightly from the carbon position, in order to reflect the presence of the hydrogen atoms. Toxvaerd has developed such a model that gives superior performance for alkanes than do the simple united atom models, particularly for simulations at high pressures [Toxvaerd 1990]. In

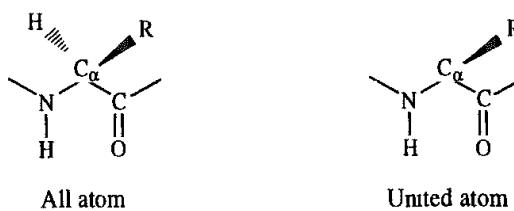


Fig 4.42 Representations of the naturally occurring amino acids

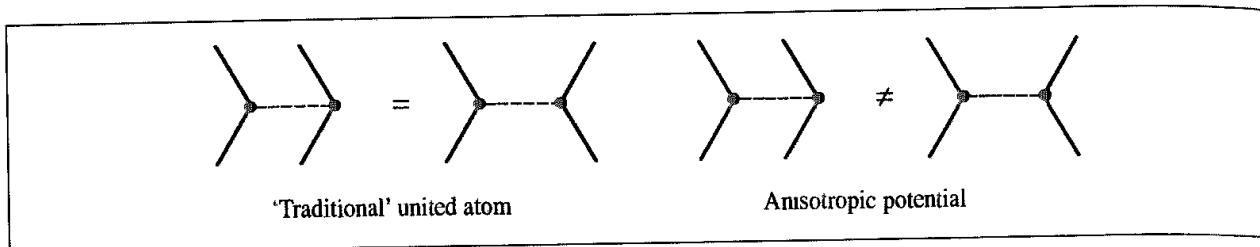


Fig. 4.43: The interaction energy between the two arrangements shown is equal in a 'traditional' united atom force field but different in the Toxvaerd anisotropic model (Figure adapted from Toxvaerd S 1990 Molecular Dynamics Calculations of the Equation of State of Alkanes The Journal of Chemical Physics 93:4290-4295)

In Toxvaerd's model the interaction sites are located at the geometrical centres of the CH₂ or CH₃ groups. The forces between these sites act on the united atom mass centre, which remains located on the carbon atom (with a mass of 14 for a CH₂ group and 15 for a CH₃ group). As the interaction site is no longer located at an atomic nucleus the forces acting on the masses are more complicated to calculate, but little additional computational expense is required. The effect of using such an anisotropic potential is nicely illustrated by the two arrangements of methylene units shown schematically in Figure 4.43. In the united atom model both arrangements would have the same energies and forces, but this is not so with the Toxvaerd anisotropic potential.

4.15.1 Other Simplified Models

In some force field models, even simpler representations are used than the united atom approach, with entire groups of atoms being modelled as single interaction points. For example, a benzene ring might be modelled as a single site with appropriately chosen parameters.

Yet other models have no obvious relationship to any ‘real’ molecule but are useful because their simplicity enables larger or more extensive calculations to be performed than would otherwise be possible. The polymer field is full of such models, as we shall discuss in Section 8.6. Another area where such models have been widely applied is in the study of liquid crystals. Liquid crystals are able to form phases that are characterised by a long-range order of the molecular orientations in at least one dimension. Many of the molecules that exhibit liquid crystalline behaviour are rod-shaped, but disc-like molecules can also form liquid crystalline phases. Some typical examples of molecules that can show such behaviour are shown in Figure 4.44. In the liquid crystalline state the rod-shaped molecules are aligned with their long axes pointing in approximately the same direction. Some very simple computer models have been used to investigate the behaviour of liquid crystals. These simple models enable large simulations to be performed on assemblies of many ‘molecules’. One example of such a simplified model is the Gay–Berne potential [Gay and Berne 1981], which models the anisotropic interaction between two particles as:

$$\nu(r_{ij}) = 4\varepsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) \left\{ \left[\frac{\sigma_0}{r_{ij} - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_0} \right]^{12} - \left[\frac{\sigma_s}{r_{ij} - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_s} \right]^6 \right\} \quad (4.90)$$

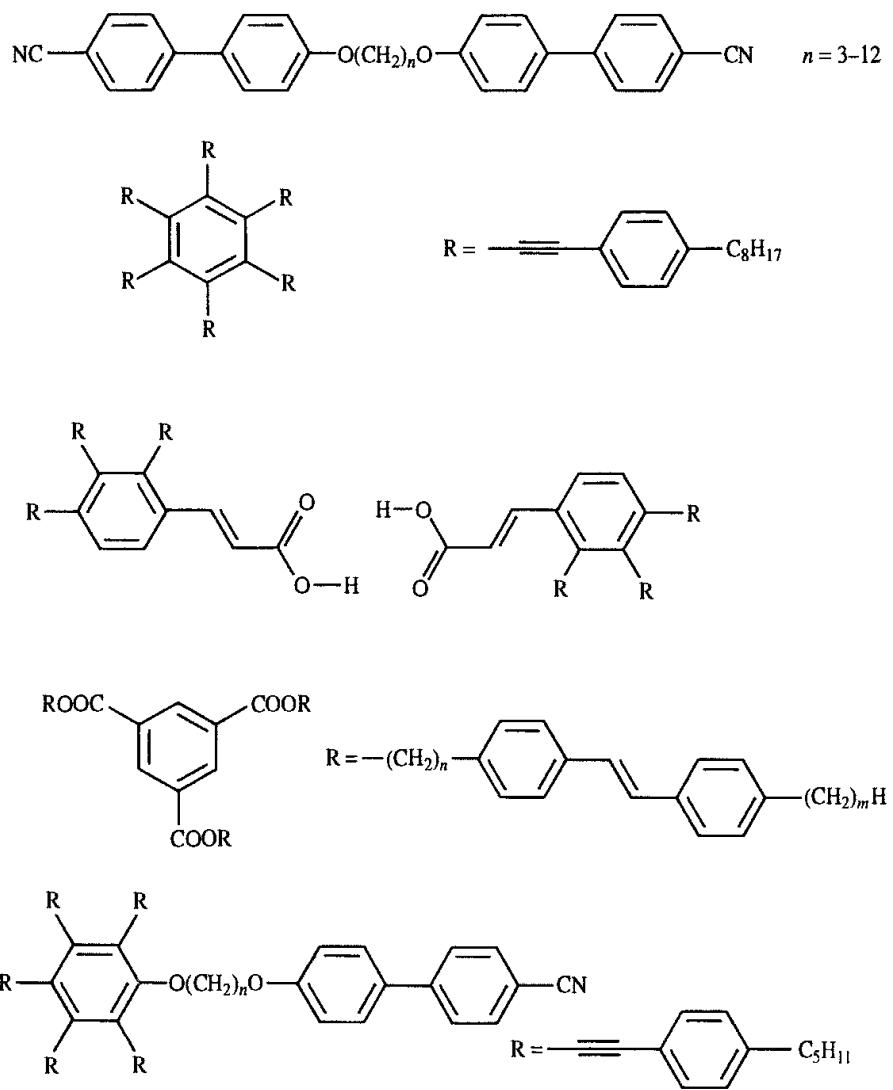


Fig. 4.44 Some typical liquid crystal molecules.

$\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$ are unit vectors that describe the orientations of the two molecules i and j and $\hat{\mathbf{r}}$ is a unit vector along the line connecting their centres (Figure 4.45). The molecules can be considered as ellipsoids which have a shape that is reflected in two size parameters, σ_s and σ_e , which are the separations at which the attractive and repulsive terms in the potential cancel for end-to-end and side-by-side arrangements respectively. These are incorporated into the potential via the parameter σ :

$$\sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) = \sigma_0 \left\{ 1 - \frac{\chi}{2} \left[\frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} + \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 + \chi(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j)} + \frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} - \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 - \chi(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j)} \right] \right\}^{-1/2} \quad (4.91)$$

where

$$\chi = \frac{(\sigma_e/\sigma_s)^2 - 1}{(\sigma_e/\sigma_s)^2 + 1} \quad (4.92)$$

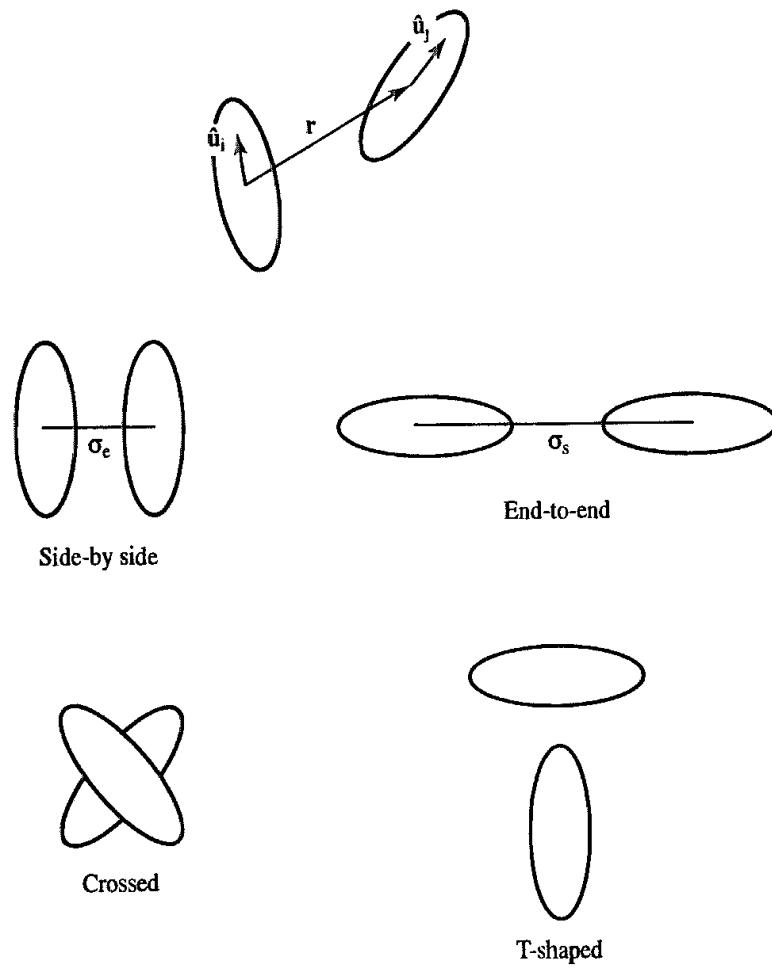


Fig. 4.45 The Gay-Berne model for liquid crystal systems and some typical arrangements.

χ is the *shape anisotropy parameter*; it is zero for spherical particles and is equal to 1 for infinitely long rods and -1 for infinitely thin discs; σ_0 is typically set equal to σ_s .

The energy term is also orientation-dependent and is written as follows:

$$\varepsilon(\hat{u}_i, \hat{u}_j, \hat{r}) = \varepsilon_0 \varepsilon'^\mu(\hat{u}_i, \hat{u}_j, \hat{r}) \varepsilon^\nu(\hat{u}_i, \hat{u}_j) \quad (4.93)$$

where

$$\begin{aligned} \varepsilon(\hat{u}_i, \hat{u}_j) &= [1 - \chi^2(\hat{u}_i \cdot \hat{u}_j)^2]^{-1/2} \\ \varepsilon'(\hat{u}_i, \hat{u}_j, \hat{r}) &= \left\{ 1 - \frac{\chi'}{2} \left[\frac{(\hat{u}_i \cdot \hat{r} + \hat{u}_j \cdot \hat{r})^2}{1 + \chi'(\hat{u}_i \cdot \hat{u}_j)} + \frac{(\hat{u}_i \cdot \hat{r} - \hat{u}_j \cdot \hat{r})^2}{1 - \chi'(\hat{u}_i \cdot \hat{u}_j)} \right] \right\} \end{aligned} \quad (4.94)$$

χ' measures the anisotropy of the attractive forces:

$$\chi' = \frac{1 - (\varepsilon_e / \varepsilon_s)^{1/\mu}}{(\varepsilon_e / \varepsilon_s)^{1/\mu} + 1} \quad (4.95)$$

ε_e is the well depth for an end-to-end arrangement of the ellipsoids when the attractive and repulsive contributions cancel, and ε_s is the corresponding well depth for the side-by-side arrangement (Figure 4.45).

The Gay-Berne potential is rather complex but is governed by a relatively small number of parameters, some of which have readily interpretable meanings. The effect of changing the parameters can be most clearly understood by considering certain orientations, such as the side-by-side, end-to-end, crossed and T-shaped structures (Figure 4.45). In the crossed structure the well depth $\varepsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}})$ and the separation $\sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}})$ are independent of χ and χ' . The ratio of the well depths for the end-to-end and side-by-side arrangements is $\varepsilon_e/\varepsilon_s$. The exponents μ and ν are considered adjustable parameters. One way to obtain values for these is to fit the Gay-Berne function to arrangements of Lennard-Jones particles. For example, Luckhurst, Stevens and Phippen determined a value of 1 for ν and a value of 2 for μ by fitting to a linear array of four Lennard-Jones centres [Luckhurst *et al.* 1990].

Depending upon the parameters chosen, simulations performed using the Gay-Berne potential show behaviour typical of liquid crystalline materials. Moreover, by modifying the potential one can determine what contributions affect the liquid crystalline properties and so help to suggest what types of molecule should be made in order to attain certain properties.

4.16 Derivatives of the Molecular Mechanics Energy Function

Many molecular modelling techniques that use force-field models require the derivatives of the energy (i.e. the force) to be calculated with respect to the coordinates. It is preferable that analytical expressions for these derivatives are available because they are more accurate and faster than numerical derivatives. A molecular mechanics energy is usually expressed in terms of a combination of internal coordinates of the system (bonds, angles, torsions, etc.) and interatomic distances (for the non-bonded interactions). The atomic positions in molecular mechanics are invariably expressed in terms of Cartesian coordinates (unlike quantum mechanics, where internal coordinates are often used). The calculation of derivatives with respect to the atomic coordinates usually requires the chain rule to be applied. For example, for an energy function that depends upon the separation between two atoms (such as the Lennard-Jones potential, Coulomb electrostatic interaction or bond-stretching term) we can write:

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (4.96)$$

$$\frac{\partial \nu}{\partial x_i} = \frac{\partial \nu}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial x_i} \quad (4.97)$$

$$\frac{\partial r_{ij}}{\partial x_i} = \frac{(x_i - x_j)}{r_{ij}} \quad (4.98)$$

Thus, for the Lennard-Jones potential:

$$\frac{\partial \nu}{\partial r_{ij}} = \frac{24\varepsilon}{r_{ij}} \left[-2\left(\frac{\sigma}{r_{ij}}\right)^{12} + \left(\frac{\sigma}{r_{ij}}\right)^6 \right] \quad (4.99)$$

The force in the x direction acting on atom i due to its interaction with atom j is given by:

$$\mathbf{f}_{x_i} = (\mathbf{x}_i - \mathbf{x}_j) \frac{24\epsilon}{r_{ij}^2} \left[2\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^6 \right] \quad (4.100)$$

Analytical expressions for the derivatives of the other terms that are commonly found in force fields are also available [Niketic and Rasmussen 1977]. Similar expressions must be derived from scratch when new functional forms are developed.

4.17 Calculating Thermodynamic Properties Using a Force Field

A molecular mechanics program will return an ‘energy value’ for any configuration or conformation of the system. This value is properly described as a ‘steric energy’ and is the energy of the system relative to a zero point that corresponds to a hypothetical molecule in which all of the bond lengths, valence angles, torsions and non-bonded separations are set to their strainless values. It is not necessary to know the actual value of the zero point to calculate the *relative* energies of different configurations or different conformations of the system.

Molecular mechanics can be used to calculate heats of formation. To do so requires the energy to form the bonds in the molecule to be added to the steric energy. These bond energies are typically obtained by fitting to experimentally determined heats of formation and are stored as empirical parameters within the force field. The accuracy with which heats of formation can be predicted with molecular mechanics is, in appropriate cases, comparable with experiment. Thus, the steric energy of a given structure may vary considerably from one force field to another, but its heat of formation should be much closer (if the force fields have been properly parametrised).

A third type of ‘energy’ that can be obtained from a molecular mechanics calculation is the ‘strain energy’. Differences in steric energy are only valid for different conformations or configurations of the same system. Strain energies enable different molecules to be compared. To determine the strain energy it is usual to define some ‘strainless’ reference point. The reference points can be chosen in many ways and so many different definitions of strain energy have been proposed in the literature. For example, Allinger and co-workers defined the reference point using a set of ‘strainless’ compounds such as the all-*trans* conformations of the straight-chain alkanes from methane to hexane. From this set of compounds it was possible to derive a set of strainless energy parameters for constituent parts of the molecules. The inherent strain energy of a hydrocarbon is then obtained by subtracting the reference ‘strainless’ energy from the actual steric energy calculated using the force field. One interesting conclusion of this study was that chair cyclohexane has an inherent strain energy due to the presence of 1,4 van der Waals interactions between the carbon atoms within the ring.

The sources of strain are often quantified by examining the different components (bonds, angles, etc.) of the force field. Such analyses can provide useful information, especially for cases such as highly strained rings. However, in many molecules the strain is distributed

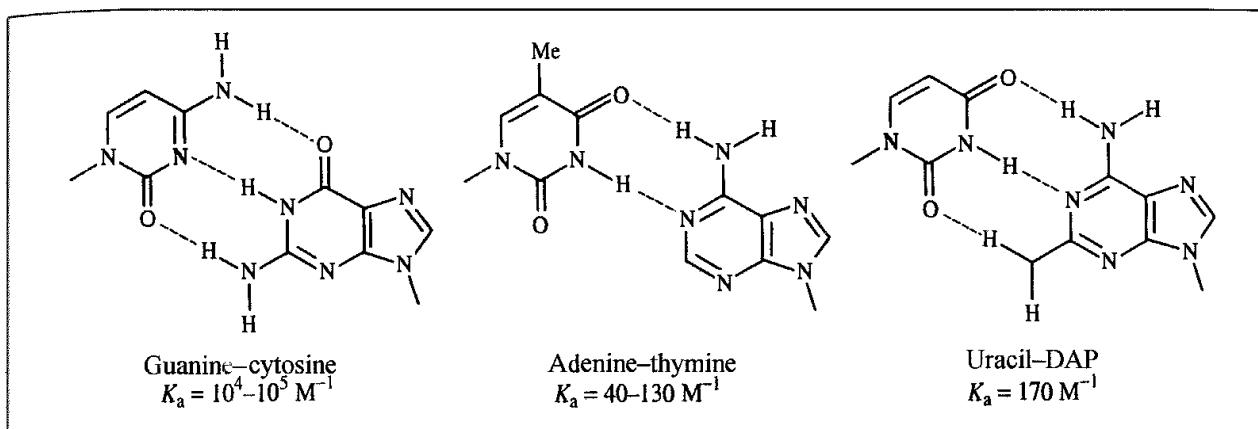


Fig. 4.46 The DNA base pairs guanine (G), cytosine (C), adenine (A) and thymine (T). The uracil-2,6-diaminopyridine pair can also form three hydrogen bonds but has a much lower association constant than G-C

among a variety of internal parameters (and in any case is force-field-dependent). For intermolecular interactions the interpretation can be easier, for the ‘interaction energy’ is simply equal to the difference between the energies of the two isolated species and the energy of the intermolecular complex. A good example of this type of calculation and the conclusions that can be drawn from it is the study by Jorgensen and Pranata [Jorgensen and Pranata 1990] of the interaction between analogues of the DNA base pairs. In the double helical structure of DNA the bases pair up adenine (A) with thymine (T) and guanine (G) with cytosine (C) (Figure 4.46).

The association constant of the G-C base pair in chloroform is between 10^4 M^{-1} and 10^5 M^{-1} whereas the association between the A-T base pair is significantly weaker, at $40\text{--}130 \text{ M}^{-1}$. One obvious reason for this difference is that there are three hydrogen bonds in the G-C base pair and only two in the A-T base pair. However, a simple hydrogen-bond count

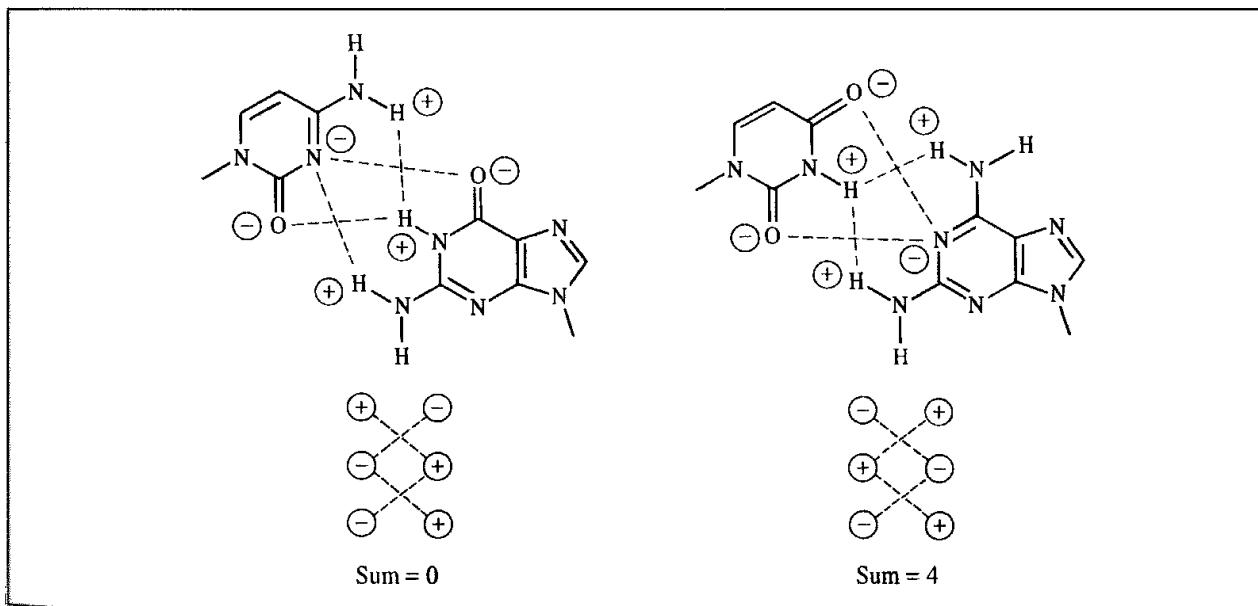


Fig. 4.47: Secondary interactions in guanine-cytosine and uracil-DAP.

does not explain all of the data, for synthetic analogues show a significant variation in their association constants, despite having three hydrogen bonds. The weak binding of the uracil-2,6-diaminopyridine (DAP) system (Figure 4.46) could be considered especially anomalous as it contains the same types of hydrogen bond as in G-C ($\text{NH}_2 \cdots \text{O}$, $\text{NH} \cdots \text{N}$, $\text{NH}_2 \cdots \text{O}$). A qualitative explanation for this phenomenon was proposed by Jorgensen and Pranata who examined the secondary interactions in these complexes. As shown in Figure 4.47, the G-C system contains two unfavourable secondary interactions and two favourable ones, an overall sum of zero. In the uracil-DAP system, all four secondary interactions are unfavourable.

4.18 Force Field Parametrisation

A force field can contain a large number of parameters, even if it is intended for calculations on only a small set of molecules. Parametrisation of a force field is not a trivial task. A significant amount of effort is required to create a new force field entirely from scratch, and even the addition of a few parameters to an existing force field in order to model a new class of molecules can be a complicated and time-consuming procedure. The performance of a force field is often particularly sensitive to just a few of the parameters (usually the non-bonded and torsional terms), so it is often sensible to spend more time optimising these parameters rather than others (such as the bond-stretching and angle-bending terms), the values of which do not greatly affect the results.

The first step is to select the data that are going to be used to guide the parametrisation process. Molecular mechanics force fields may be used to determine a variety of structurally related properties and the parametrisation data should be chosen accordingly. The geometries and relative conformational energies of certain key molecules are usually included in the data set. It is increasingly common to include vibrational frequencies in the parametrisation; these are usually more difficult to reproduce but the incorporation of appropriate cross terms can often help. Some force fields are parametrised to reproduce thermodynamic properties using computer simulation techniques. The OPLS (optimised parameters for liquid simulations [Jorgensen and Tirado-Reeves 1988]) parameters have been obtained in this way.

Unfortunately, experimental data may be non-existent or difficult to obtain for particular classes of molecules. Quantum mechanics calculations are thus increasingly used to provide the data for the parametrisation of molecular mechanics force fields. This is an important development because it greatly extends the range of chemical systems that can be treated using the force-field approach. *Ab initio* calculations are able to reproduce experimental results for small representative systems. Clearly, one should be careful to properly validate a force field derived in such a way by testing against experimental data if at all possible.

Once a functional form for the force field has been chosen and the data to be used in the parametrisation identified, there are then two basic methods that can be used to actually obtain the parameters. The first approach is ‘parametrisation by trial and error’, in which

the parameters are gradually refined to give better and better fits to the data. It is difficult to simultaneously modify a large number of parameters in such a strategy and so it is usual to perform the parametrisation in stages. It is important to remember that there is some coupling between all of the degrees of freedom and so for the most sensitive work none of the parameters can truly be taken in isolation. Parameters for the hard degrees of freedom (bond stretching and angle bending) can, however, often be treated separately from the others (indeed the bond and angle parameters are often transferred from one force field to another without modification). By contrast, the soft degrees of freedom (non-bonded and torsional contributions) are closely coupled and can significantly influence each other. One protocol that can be quite successful is to first establish a series of van der Waals parameters. The electrostatic model is then determined (e.g. by electrostatic potential fitting). Finally, the torsional potentials are determined by ensuring that the torsional barriers are reproduced together with the relative energies of the different conformations. Of course, it may be necessary to modify any of the parameters at any stage should the results be inadequate and so parametrisation is invariably an iterative procedure.

As experimental information on torsional barriers is often sparse or non-existent, quantum mechanical calculations are widely used to determine torsional potentials. The general strategy is as follows. First, a molecular fragment that adequately represents the rotatable bond of interest and its immediate environment is chosen. A series of structures are then generated by rotating about the bond and their energies determined using quantum mechanics. The torsional potential is then fitted to reproduce the energy curve, in conjunction with the van der Waals potential and partial charges. This procedure can be illustrated using the study of Pranata and Jorgensen who wanted to perform some calculations on FK506, a potent immunosuppressant (Figure 4.48) [Pranata and Jorgensen 1991]. FK506 contains a ketoamide functionality that has a *trans* conformation when the molecule is bound to its receptor but which is *cis* in the crystal structure of isolated FK506. NMR experiments suggested that the molecule adopts both *cis* and *trans* conformations in solution. This part of the molecule is clearly implicated in its function and so it was considered important

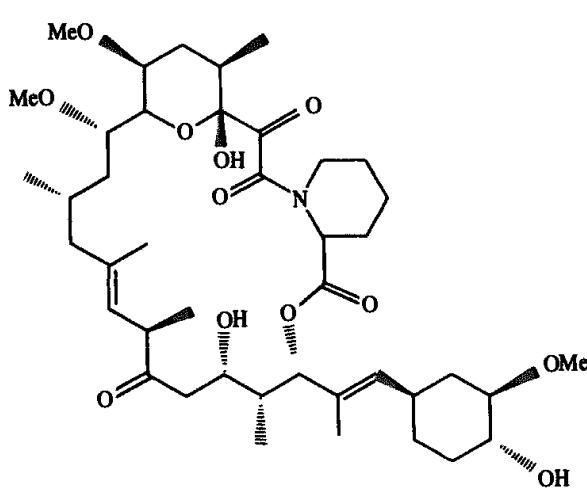


Fig. 4.48 The immunosuppressant FK506

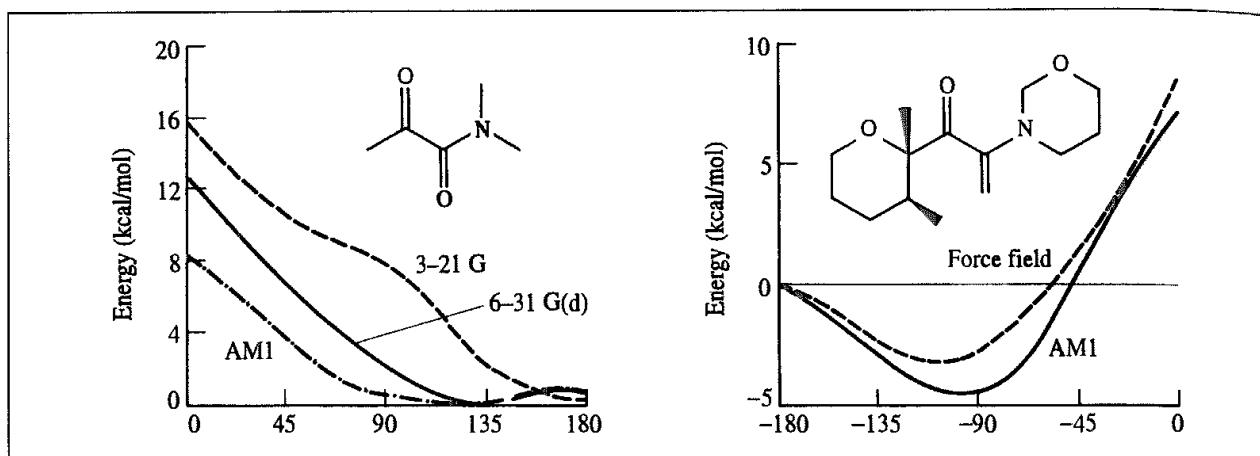


Fig 4.49 Fragments used to derive and evaluate parameters for the ketoamide functionality in FK506 (Figure redrawn from J Pranata and W L Jorgensen 1991 Computational Studies on FK506. Conformational Search and Molecular Dynamics Simulations in Water The Journal of the American Chemical Society 113:9483–9493)

to correctly model the torsional potential about this bond. Pranata and Jorgensen intended to use the AMBER force field for their calculations but the force field contained no parameters for this link.

Molecular orbital calculations were performed on *N,N*-dimethyl- α -ketopropanamide (Figure 4.49, left), which was chosen as an appropriate model system. Semi-empirical calculations using AM1 and *ab initio* calculations using a 6-31G(d) basis set suggested that the minimum energy conformation corresponded to a torsion angle of 124° and 135°, respectively, with the *anti* conformation being slightly higher in energy (~0.7 kcal/mol). However, an analogous calculation using the 3-21G basis set did predict that the *anti* conformation was at a minimum (Figure 4.49). Crystal structures of compounds containing this fragment revealed that an orthogonal structure was commonly encountered. Torsional parameters were then fitted to the 6-31G(d) potential and evaluated by calculating an energetic profile for rotation in a larger fragment of the FK506 molecule using the force field and comparing it with that obtained using AM1 (Figure 4.49, right).

An alternative approach to parametrisation, pioneered by Lifson and co-workers in the development of their ‘consistent’ force fields, is to use least-squares fitting to determine the set of parameters that gives the optimal fit to the data [Lifson and Warshel 1968]. Again, the first step is to choose a set of experimental data that one wishes the force field to reproduce (or calculate using quantum mechanics, if appropriate). Warshel and Lifson used thermodynamic data, equilibrium conformations and vibrational frequencies. The ‘error’ for a given set of parameters equals the sum of squares of the differences between the observed and calculated values for the set of properties. The objective is to change the force field parameters to minimise the error. This is done by assuming that the properties can be related to the force field by a Taylor series expansion:

$$\Delta\mathbf{y}(\mathbf{x} + \delta\mathbf{x}) = \Delta\mathbf{y}(\mathbf{x}) + \mathbf{Z}\delta\mathbf{x} + \dots \quad (4.101)$$

$\Delta\mathbf{y}$ is a vector of the differences between the calculated and experimental data and is a vector whose components are the force field parameters. \mathbf{Z} is a matrix whose elements are the

derivatives of each property with respect to each of the parameters, $\partial x/\partial y$. An iterative procedure is used to minimise the sum of squares of the differences, Δy^2 . The method is easily modified to enable various weighting factors to be assigned to the different pieces of experimental data, so that (for example) the thermodynamic data could be given greater importance than the vibrational frequencies.

A well-known application of the least-squares approach to the optimisation of a force field was performed by Hagler, Huler and Lifson, who derived a force field for peptides by fitting to crystal data of a variety of appropriate compounds [Hagler *et al.* 1977; Hagler and Lifson 1974]. A key result of their work was that no explicit hydrogen bond term was required to model the hydrogen-bonding interactions, but that a combination of appropriate electrostatic and van der Waals models was sufficient. A group led by Hagler more recently developed a force field based upon the results of *ab initio* quantum mechanics calculations on small molecules, again using least-squares fitting [Maple *et al.* 1988]. The quantum mechanics calculations were performed not only on small molecules at equilibrium geometries but also on structures that were distorted from equilibrium. For each geometry the energy was calculated together with the first and second derivatives of the energy. This provided a wealth of data for the subsequent fitting procedure. This research has resulted in many new algorithms for the derivation of force-field parameters and has also challenged some of the assumptions about the development and functional form of force fields. One feature of the resulting force field, named CFF (standing for consistent force field), is that it contains rather more cross terms than other force fields. This can be ascribed to the objective of accurately reproducing vibrational spectra.

4.19 Transferability of Force Field Parameters

The range of systems that have been studied by force field methods is extremely varied. Some force fields have been developed to study just one atomic or molecular species under a wider range of conditions. For example, the chlorine model of Rodger, Stone and Tildesley [Rodger *et al.* 1988] can be used to study the solid, liquid and gaseous phases. This is an anisotropic site model, in which the interaction between a pair of sites on two molecules depends not only upon the separation between the sites (as in an isotropic model such as the Lennard-Jones model) but also upon the orientation of the site-site vector with respect to the bond vectors of the two molecules. The model includes an electrostatic component which contains dipole-dipole, dipole-quadrupole and quadrupole-quadrupole terms, and the van der Waals contribution is modelled using a Buckingham-like function.

Other force fields are designed for use with specific classes of molecules; we have already encountered the AMBER force field, which is designed for calculations on proteins and nucleic acids. Yet other force fields are intended to be applied to a wide range of molecules, and indeed some force fields are designed to model the entire periodic table. Intuitively, one might expect a 'specialised' force field to perform better than a 'general' force field, and while this is certainly true for the best of the specialised force fields, a good general force field can often outperform a poor specific force field.

The ability to transfer parameters from one molecule to another is crucial for any force field. Without it, the task of parametrisation would be impossible, because so many parameters would be required, and the force field would have no predictive ability. Transferability has a number of important consequences for the development and application of force fields. The problem of transferability is often first encountered when a molecular mechanics program fails to run because parameters are missing for the molecule being studied. One must somehow find values for the missing parameters. Some programs automatically 'guess' force field parameters; it is wise to check these assignments as they may be suspect. For the developer of a force field, a compromise must often be found between a complex functional form and a large number of atom types. It is also important to try to ensure that the errors in the force field are balanced, in the sense that it would be silly to spend a lot of time getting (say) the bond-stretching terms just right, if the van der Waals parameters give rise to large errors.

An alternative to 'guessing' parameters (which, if done properly, can sometimes give quite reasonable results) is to construct the force field in such a way that the parameters can be derived from atomic properties. This is particularly pertinent to those force fields which are designed to be used on a very wide range of elements and atom types, such as the Universal Force Field [Rappé *et al.* 1992]. This force field is claimed to model the entire periodic table and as such it would probably be impossible to derive individual parameters for each of the terms; indeed, the data required for such an exercise does not exist for many cases. Thus the UFF has a set of atom types which are characterised by atomic number, hybridisation and formal oxidation state. Reference bond lengths are initially set equal to the sum of the two relevant atomic bond radii and then corrected for bond order and the relative electronegativities of the two atoms. Bond force constants are obtained from Badger's rules, under which the force constant is proportional to the product of the 'effective atomic charges' for the two atoms and inversely proportional to the cube of the interatomic distance:

$$k_{ij} \propto \frac{q_i^* q_j^*}{r_{ij}^3} \quad (4.102)$$

The effective atomic charges are either obtained by fitting to data on diatomic molecules (where it exists) or by interpolation or extrapolation from this fit.

Transferability can be helped by using the same parameters for as wide a range of situations as possible. The non-bonded terms are particularly problematic in this regard; it would, in principle, be necessary to have parameters for the non-bonded interactions between all possible pairs of atom types. This would give rise to a very large number of parameters. It is therefore commonly assumed that the same set of van der Waals parameters can be used for most, if not all, atoms of the same element. For example, all carbon atoms (sp^3 , sp^2 , sp , etc.) would be treated with the same set of van der Waals parameters, all nitrogens by a common set, and so on. The torsional terms may also be generalised, so that the torsional parameters depend solely upon the atom types of the two atoms that form the central bond, rather than on all four atoms that comprise the torsion angle, as described in Section 4.5 for the AMBER force field.

4.20 The Treatment of Delocalised π Systems

The bonds in conjugated π systems are often of different lengths. For example, the central bond in butadiene is approximately 1.47 Å long, but the two terminal CH=CH₂ bonds are approximately 1.34 Å. If butadiene is modelled using a force field in which all four carbon atoms are assigned the same atom type (e.g. 'carbon sp²') then each bond will be assigned the same bonding parameters and in the equilibrium structure all carbon–carbon bonds will be almost identical in length. A similar situation arises for aromatic systems. For example, not all the bonds in naphthalene are of equal length (unlike benzene). The bond lengths in a delocalised π system depend upon the bond orders; the higher the bond order, the shorter the bond.

In some cases it may be possible to circumvent this problem by creating a model specific to the conjugated system. For butadiene the central carbon–carbon bond of the π system could be treated in a different manner to the two terminal bonds, for example by using one atom type for the –CH= carbon atoms and one for the =CH₂ carbon atoms in butadiene. This approach might be acceptable if we wanted to perform an extensive series of calculations on substituted butadienes, but it does compromise the transferability of the force field parameters. An alternative is to incorporate a molecular orbital calculation into the force field. Two variants on this theme have been developed. In one approach, the π and σ systems are treated separately [Warshel and Karplus 1972; Warshel and Lappicirella 1981]. For a given geometry, a self-consistent field quantum mechanical calculation is performed on the π system, typically with an appropriate semi-empirical theory. Molecular mechanics is simultaneously applied to the σ system. The energies of the quantum mechanical and molecular mechanical calculations are added together, and the geometry is modified to minimise this combined energy. A obvious assumption inherent in this approach is that the π and σ systems can be separated, which may be difficult to justify when deviations from planarity are present. Nevertheless, the approach has been extended to include those containing conjugated nitrogen and oxygen atoms, which has enabled the study of the properties of not only the ground states of some important biological chromophores (such as porphyrins) but also their excited states [Warshel and Lappicirella 1981].

An alternative approach is exemplified by the MM2/MM3/MM4 family of programs. First, a molecular orbital calculation is performed on the π system. If the initial conformation of the system is non-planar the calculation is performed on the equivalent planar system. The force field parameters are then modified according to the quantum mechanical bond orders. In MMP2 (the name given to the special version of MM2 which incorporated these features) these parameters are the force constant for the bonds in the π system, the reference bond lengths and the torsional barriers [Sprague *et al* 1987; Allinger and Sprague 1973]. The system is then subjected to the usual molecular mechanics treatment using the new force field parameters. A linear relationship between the stretching constants and the bond orders, and between the reference bond lengths and the bond orders was found to give good results. Initially, the torsional barriers were assumed to be proportional to the square of the bond orders, but this relationship was modified slightly in subsequent versions

of the program. Thus in MM4 the V_2 and V_3 terms become:

$$V_2 = [A + p_{ij}^{\omega=0} \beta_{ij}] V_2^0 \quad (4.103)$$

$$V_3 = K_{V_3} [1 - p_{ij}(\omega)] V_3^0 \quad (4.104)$$

In Equation (4.103) p_{ij} is the bond order about the central bond $i-j$ of the torsion angle calculated for a torsion angle of zero and β_{ij} is the resonance integral from the molecular orbital calculation. The parameter A has a value of -0.09 and so the V_2 term is lower for those conjugated bonds with a lower bond order. In Equation (4.104) p_{ij} is now the bond order for the bond $i-j$ calculated for the torsion angle ω . K_{V_3} equals 1.25 and so V_3 increases with decreasing bond order. A bond with a lower bond order (and so a lower V_2 and a higher V_3) is thus more likely to deviate from planarity.

4.21 Force Fields for Inorganic Molecules

It may come as a surprise to many readers to learn that the earliest force field calculations on inorganic molecules were reported at much the same time as the first calculations on organic systems. For example, Corey and Bailar described the use of empirical force field calculations on octahedral complexes of cobalt in 1959 [Corey and Bailar 1959]. The range of metal-containing systems that can be considered by force field methods has steadily expanded since then. Moreover, many systems of commercial interest contain metals or other elements not usually found in 'organic' or 'biochemical' systems.

Some inorganic systems (such as certain coordination complexes) are little different to organic systems from a force field point of view; the bonding can be represented in a similar way and many of the force field parameters originally developed for organic systems can be transferred without modification. However, inorganic molecules do have certain properties which makes them more difficult to model than their organic counterparts. Perhaps the two most striking properties are the much wider range of geometries and the presence of highly delocalised bonds. Thus inorganic molecules include square planar and sawhorse (e.g. SF_4) shapes for four coordination and T-shaped for three coordination. Coordination numbers higher than four are also possible, with five (square pyramidal, trigonal bipyramidal) and six (octahedral and trigonal prismatic) being particularly common. To model such systems using conventional organic force fields would often be problematic because their geometries do not have a high degree of symmetry. For example, in a trigonal bipyramidal there are in principle three different types of bond angle subtended at the central atom (90° , 120° and 180°). Moreover, in such systems the atoms are often equivalent (interchanging them gives the same structure back). However, if these atoms are assigned different force field parameters then this equivalence is not reproduced by the calculation. At least in these cases there is an obvious localised bonding scheme that can be applied; this is often not possible with organometallic molecules. For example, how should the bonding in ferrocene be represented in a force field calculation? Is there a bond between the iron and each of the carbon atoms in the two cyclopentadienyl rings? Is there a 'bond' from the iron to the centre of each of the rings? A yet further complication is that significant deviations from ideal geometries are often observed due to electronic effects such as the Jahn-Teller effect.

Whilst there is no universal solution to these problems within the context of a single force field similar to those used in organic chemistry, for certain situations it is possible to use an organic-like force field with only relatively small modifications. For obvious reasons those complexes with a high degree of symmetry are most amenable to such a treatment. Thus octahedral and square planar complexes are the simplest to model because of their symmetry (in addition to the geometries common in organic chemistry). However, even these have two types of equilibrium angle (180° and 90°). The situation can be much more complicated for the other geometries or for structures where the geometry about the metal is a distortion of a regular arrangement. A Urey-Bradley treatment of the bonding about the metal can often be quite successful in achieving the correct geometries. Here, there are no angle-bending terms at the metal but terms due to pairs of atoms bonded to the metal.

It is much more difficult to use such a force field to model metal π systems, where the bonding between the metal and the ligand is not easily represented by a conventional bonding picture. As we have discussed, metal atoms can adopt a wide range of geometries in π complexes, which are often significantly distorted from regular structures. Nevertheless, force fields have been developed which can cope with such systems, as well as being able to model more traditional systems such as organic compounds. These force fields often use a rather different functional form from Equation (4.1) and the parameters are obtained in a different way. One distinctive feature of both the Universal Force Field and the SHAPES force field developed by Landis and co-workers [Allured *et al.* 1991; Cleveland and Landis 1996] is the way in which angle bending is treated. The harmonic potential that is commonly employed in standard force fields is inappropriate to model the distortion of systems as the angle approaches 180° . UFF [Rappé *et al.* 1993] uses a cosine Fourier series for each angle ABC:

$$\nu(\theta) = K_{ABC} \sum_{n=0}^m C_n \cos n\theta \quad (4.105)$$

The coefficients C_n are chosen to ensure that the function has a minimum at the appropriate reference bond angle. For linear, trigonal, square planar and octahedral coordination, Fourier series with just two terms are used with a C_0 term and a term for $n = 1, 2, 3$ or 4, respectively:

$$\nu(\theta) = K_{ABC}[1 - \cos(n\theta)] \quad (4.106)$$

Thus, for example, if $n = 4$ then the function has minima at both 90° and 180° as required for octahedral geometries. The general case is exemplified by the H-O-H angle in water, where it is desired to have a minimum in the energy at an angle of 104.5° . Moreover, at this angle (θ_0) the second derivative of the energy equals the force constant. If in addition it is required that the energy is a maximum at 180° the following expression results:

$$\nu(\theta) = K_{ABC}[C_0 + C_1 \cos(\theta) + C_2(\cos 2\theta)] \quad (4.107)$$

The three coefficients are defined as:

$$C_2 = \frac{1}{4 \sin^2(\theta_0)}; \quad C_1 = -4C_2 \cos(\theta_0); \quad C_0 = C_2[2 \cos^2(\theta_0) + 1] \quad (4.108)$$

The SHAPES angle-bending term is very similar:

$$\nu(\theta) = K_{ABC} \sum_{n=0}^m [1 + \cos(n\theta - \delta)] \quad (4.109)$$

δ is the phase shift. Landis subsequently developed a formulation (called VALBOND) for the angle-bending term that is based on valence bond theory and which can produce results that compare well with *ab initio* calculations [Landis *et al.* 1995, 1998]. For example, using just one set of C–H parameters the H–C–H bond angles in ethene, formaldehyde and both singlet and triplet carbene match closely those found experimentally. One key practical advantage of this method is that it is not necessary to define equilibrium bond angles.

4.22 Force Fields for Solid-state Systems

Empirical potential models are widely used to study the solid state, complementing the quantum mechanical approaches we discussed in Chapter 3. One important difference between solid-state materials and ‘organic’ molecules (and indeed, some inorganic complexes) is that whilst the latter can generally be described using a localised bond model this is not always the case for the former. As a consequence, molecular mechanics approaches of the kind we have discussed so far in this chapter can be applied successfully only to certain types of material. Ionic and metallic systems especially require an alternative approach. Perhaps the key difference between solid-state materials and isolated molecules is the way in which the electrostatic terms are considered. As we shall see in Sections 6.7 and 6.8 it is common to truncate such interactions at some cutoff distance. However, solid-state modelling is concerned with materials that have long-range order; moreover, they often contain highly charged species. This means that the use of cutoffs can have a particularly detrimental effect, necessitating the use of special techniques such as the Ewald summation that enable more accurate interaction energies to be calculated. First, however, we shall consider the treatment of covalent systems which are amenable to the ‘organic’ style of molecular mechanics force field treatment, as exemplified by the study of zeolites.

4.22.1 Covalent Solids: Zeolites

Zeolites are materials generally composed of silicon, aluminium, oxygen and a metal cation or proton. They have a multitude of commercial uses including catalysis and separation (e.g. they are used in oil refining to separate linear and branched alkanes). Many of these important properties are a consequence of the presence within the zeolite of channels of molecular dimensions. It is therefore natural that molecular modelling techniques should be used to investigate the intrinsic properties of such materials and the way in which they interact with adsorbates.

The size of many zeolite systems means that considerable computational resources may be required for the calculation. In some cases therefore, such as the study of adsorption

processes, the zeolite is kept rigid and attention is concentrated on the intermolecular interactions between the zeolite and the adsorbate. This is often done using a combination of van der Waals and electrostatic terms; a Lennard-Jones potential may be used for the van der Waals component, but a Buckingham-like potential is often preferred. Electrostatic interactions can be very important for zeolites. However, the partial charges used in the various published force fields can vary enormously (from $0.4e$ to as much as $1.9e$ for the silicon atoms in silicates).

It is obviously an approximation to keep the zeolite rigid, and in more complex models the structure can vary. Many of the force fields that have been developed to model zeolites are very similar to the valence force fields used for organic and biological molecules, typically containing bond-stretching, angle-bending and torsional terms in addition to the non-bonded interactions. One important consideration when modelling zeolites is that very little energy is required to deform the Si–O–Si bond over an extremely wide range (at least 120° to 180°). This is shown in Figure 4.50, which shows the results of *ab initio* calculations using a 3-21G* basis set for $\text{H}_3\text{SiOSiH}_3$. The Fourier series expansions used by the UFF and SHAPES force fields for the angle-bending terms are designed to cope with such angular variation; Nicholas, Hopfinger, Trouw and Iton suggested the following quartic potential as an alternative specifically for the Si–O–Si angle [Nicholas *et al.* 1991]:

$$\nu(\theta) = \frac{k_1}{2}(\theta - \theta_0)^2 + \frac{k_2}{2}(\theta - \theta_0)^3 + \frac{k_3}{2}(\theta - \theta_0)^4 \quad (4.110)$$

With the correct choice of the parameters k_i and θ_0 the *ab initio* data in Figure 4.50 could be reproduced very well. In this force field a Urey-Bradley term was also included between the silicon atoms in such angles to model the lengthening of the Si–O bond as the angle decreased.

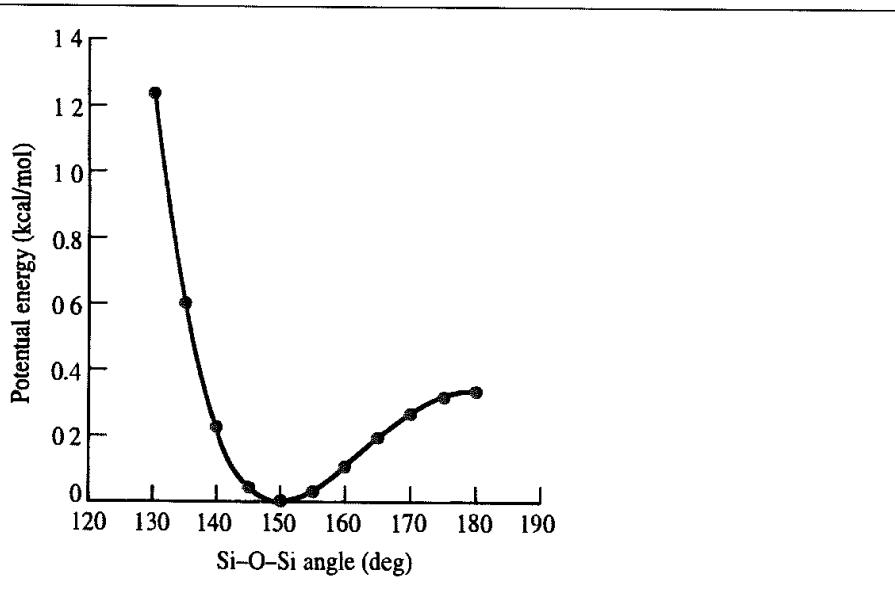


Fig 4.50. Variation in energy with the Si–O–Si angle (Figure redrawn from Grigoras S and T H Lane 1988 Molecular Parameters for Organosilicon Compounds Calculated from Ab Initio Computations Journal of Computational Chemistry 9,25–39.)

4.22.2 Ionic Solids

The covalent approach is rarely appropriate for ionic and polar solids such as oxides and halides. The usual starting point for studying such systems is to write the potential as a series expansion of pairwise, three-body, etc., terms:

$$\mathcal{V} = \mathcal{V}_0 + \sum_{i=1}^N \sum_{j=i+1}^N v_{ij}(r) + \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N v_{ijk}(r) + \dots \quad (4.111)$$

One of the oldest of such models is due to Born [Born 1920], who restricted the series to pairwise terms, which were in turn divided into long-range Coulomb interactions and short-range repulsive forces. If an inverse power law is used for the repulsive term the potential energy is thus:

$$\mathcal{V} = \sum_{i=1}^N \sum_{j=i+1}^N \left(\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \frac{A}{r_{ij}^n} \right) \quad (4.112)$$

The simplest way to apply such an equation is to assume that the charges q are equal to the oxidation states of the relevant species and that the repulsive potential only acts between nearest neighbours (though in common with many solid-state calculations the long-range ionic interaction is generally calculated for all possible interactions using an approach such as the Ewald sum, Section 6.8). This only leaves the two parameters A and n whose determination in principle requires only two pieces of experimental data (though the values obtained may vary quite considerably depending upon which data is chosen). An obvious extension of the simple form of Equation (4.112) is to model the short-range interactions by an alternative functional form; the Buckingham potential is commonly employed.

For a simple material such as sodium chloride the oxidation state assumption is a reasonable one. However, for other systems this is not necessarily the case. Various methods have been proposed for determining appropriate sets of non-integral charges. One strategy is to examine the distribution of charge within the material, as can be obtained from high-resolution X-ray experiments. However, there is no unique way to partition the charge unless there is zero bonding overlap between the ions. The atoms-in-molecules approach (see Section 2.7.7) may be a good way to do this but this is not the only option. It is worth mentioning that one advantage of the formal charge approach is that it can facilitate the transferability of potentials from one material to another whilst still maintaining charge neutrality.

The Born model with integral or partial charges assumes that the ions have zero polarisability. This is reasonable for small cations such as Li^+ or Mg^{2+} but can introduce significant errors for other systems. One property that clearly demonstrates this is the high-frequency dielectric constant. At a suitably high frequency only the electrons can keep up with the external field and the dielectric constant is given by the Clausius–Mosotti relationship:

$$\frac{(\epsilon_r - 1)}{(\epsilon_r + 2)} = \frac{4\pi}{3V_m} \sum_{i=1}^N \alpha_i \quad (4.113)$$

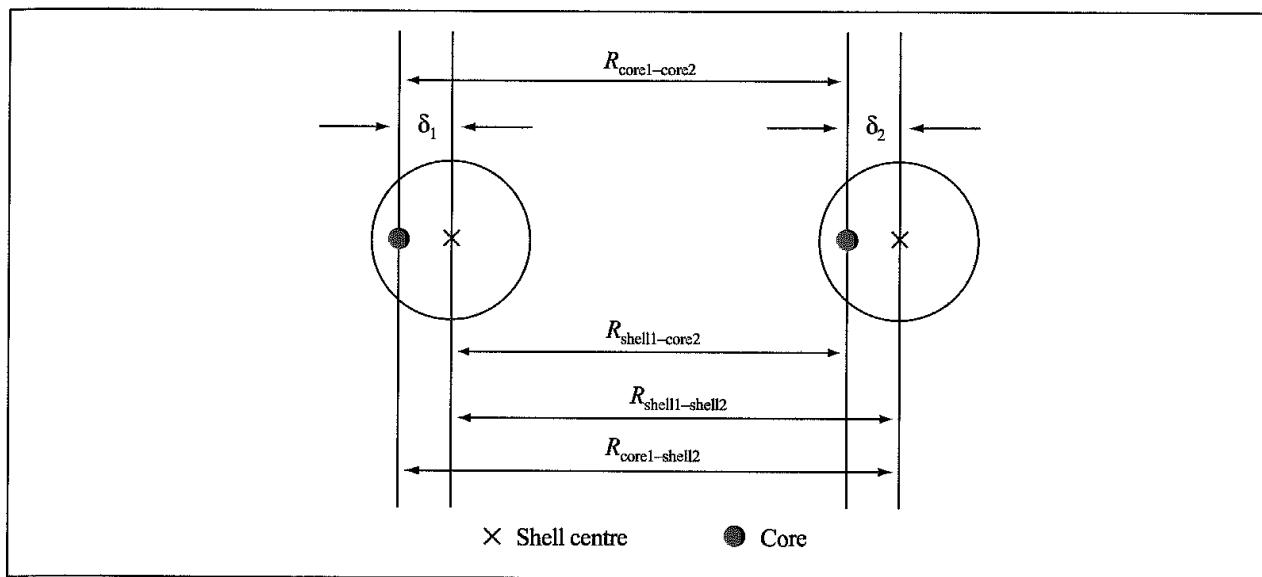


Fig 4.51 The Dick-Overhauser shell model

ϵ_r is the relative permittivity, V_m is the molar volume and α_i is the polarisability of the i th ion with the sum being over the N ions. If the ions were not polarisable then ϵ_r would have a value of 1. As we have seen, one way to incorporate polarisation is to assign a point polarisability to each ion. However, this model does not often give good results, at least for certain properties. This is because it fails to account for the coupling between polarisation and short-range repulsion effects. Thus polarisation causes distortions in the distribution of the valence electrons, and short-range repulsion is itself a consequence of the overlap between such electrons. The overall effect of short-range repulsion is to reduce polarisation effects. One model that can take this coupling into account is the shell model of Dick and Overhauser [Dick and Overhauser 1958] (Figure 4.51). In this model the ion is represented by a massive core linked to a massless shell by a harmonic spring. Both the core and the shell have charges associated with them. In an electric field the shell retains its charge but moves with respect to the core. The polarisability of an isolated ion in this model is proportional to Y^2/k where k is the spring constant of the harmonic spring and Y is the charge on the shell. The electrostatic interaction energy equals the sum over all ions and shells, not counting any interaction between an ion and its own shell. Although it is appealing to assume that the shells somehow play the role of the valence electrons this is probably an over-interpretation if only due to the fact that the shell charges, Y , do not necessarily assume small negative values.

Three-body and higher terms are sometimes incorporated into solid-state potentials. The Axilrod-Teller term is the most obvious way to achieve this. For systems such as the alkali halides this makes a small contribution to the total energy. Other approaches involve the use of terms equivalent to the harmonic angle-bending terms in valence force fields; these have the advantage of simplicity but, as we have already discussed, are only really appropriate for small deviations from the equilibrium bond angle. Nevertheless, it can make a significant difference to the quality of the results in some cases.

As for molecular systems, the parameters used to study the solid state can be derived using both experimental and theoretical data. There is a long tradition of using quantum mechanical calculations to extract such potentials. Whereas it is now common for the sophisticated Hartree–Fock and density functional theory approaches to be used for such parameter derivations, an approach called electron gas theory (a crude version of density functional theory) played a significant historical role and is still used [Allan and Mackrodt 1994]. One example of the way in which *ab initio* quantum mechanical calculations can play a role in this process is provided by the derivation of a potential model for $\alpha\text{-Al}_2\text{O}_3$ [Gale *et al.* 1992]. Previous attempts to derive empirical potentials for this material (using a shell model combined with a Buckingham potential) were not entirely successful; in particular these did not correctly predict that the corundum structure should have the lowest energy. One interesting feature of these earlier parameterisations was the great variation in the core and shell charges; for example, in one of the models the aluminium core and shell charges were 1.617 and 1.383 respectively; in another they were 10.6063 and –8.0563. A feature of the periodic Hartree–Fock calculations (see Section 3.8.3) was the use of distorted structures to provide more information on the nature of the energy surface, which was found to give better results.

4.23 Empirical Potentials for Metals and Semiconductors

Perhaps the most important consideration when discussing the development and use of empirical potentials for studying atomic solids is that pairwise potential models are often not very suitable. The performance of pairwise potential models can be bad for transition metals and even worse for semiconductors! There are a number of reasons why this is so, many of which are due to the fundamental behaviour of pairwise potentials for certain experimental properties. The most oft-quoted properties are as follows:

1. The ratio between the cohesive energy and the melting temperature, $E_c/k_B T$. The cohesive energy is the energy cost of removing an atom from within the solid matrix. This ratio is observed to be approximately 30 in metals but about 10 in pairwise systems.
2. The ratio between the vacancy formation energy and the cohesive energy, E_v/E_c . This ratio is between $\frac{1}{4}$ and $\frac{1}{3}$ in metals but closer to unity in two-body systems (exactly 1 if the structure is not permitted to relax). This can be understood as follows. Suppose each atom in a solid has Z neighbours. If one of the atoms is removed then the coordination of the surrounding Z atoms will fall to $Z - 1$. Using a pairwise energy model the vacancy formation energy is thus Z times the atom–atom bond energy. The cohesive energy is the energy to reduce the coordination of an atom from Z to zero and so would also equal Z times the atom–atom bond energy. The energy change for both of these processes is thus equal for the pairwise model.
3. The ratio between the elastic constants C_{12}/C_{44} . Elastic constants will be discussed in Section 5.10; for a cubic solid there are three distinct values, which are labelled C_{11} , C_{12} and C_{44} . For a two-body system the ratio is exactly 1 (this is known as the Cauchy relationship). For metals and oxides deviation from unity is common, gold has a particularly high value, which is indicative of its high malleability.

4. The surface properties of metals are such that the surface tends to relax inwards but systems described by two-body interactions tend to relax outwards.

The main reason for the failure of pairwise potentials is that they are unable to deal simultaneously with both surface and bulk environments. Thus on the surface there are generally fewer bonds, but these tend to be stronger than in the bulk, where there are more, but weaker, bonds. Several many-body potentials have been devised to try to address this problem. Many of these potentials have a similar, sometimes mathematically equivalent, functional form. This reflects their common origins in some form of quantum mechanical description of bonding. However, they differ in their underlying approach, the degree to which they conform to these quantum mechanical origins and the way in which they are parametrised. Here we will outline various models: the Finnis-Sinclair model (and the Sutton-Chen extension), the embedded-atom model, the Stillinger-Weber model and the Tersoff model.

The origins of the Finnis-Sinclair potential [Finnis and Sinclair 1984] lie in the density of states and the *moments theorem*. Recall that the density of states $D(E)$ (see Section 3.8.5) describes the distribution of electronic states in the system. $D(E)$ gives the number of states between E and $E + \delta E$. Such a distribution can be described in terms of its *moments*. The moments are usually defined relative to the energy of the atomic orbital from which the molecular orbitals are formed. The m th moment, μ^m , is given by:

$$\mu^m = \sum_n (E - E_{\text{atomic}})^m D(E) \quad (4.114)$$

The summation runs over the molecular orbitals or bonds. The first moment is the mean of the distribution. If the moments are defined relative to the atomic orbital energy then this first moment will be zero. The second moment (the sum of the squares of the deviations) is the width of the distribution (the variance). The third moment describes how skewed the distribution is about the mean. If all the moments are known then the distribution can be completely characterised. Of these various moments one would expect the second to be most related to the binding energy, as this indicates how much the energy levels in the solid differ from those in the atom. Indeed, a high correlation is found to exist between the binding energy and the square root of the second moment. Armed with this relationship it would be possible to predict the binding energy for perfect lattices where the atomic environments were identical. However, a more useful model is one based on a local atomic environment ('real' materials contain features such as surfaces and defects). This requires a local density of states to be defined for each atom, $d_i(E)$, where the contribution of each molecular orbital is weighted by the amount of the orbital on the atom. In a linear combination of atomic orbitals (LCAO) model this weight is the sum of the squares of the basis set coefficients for those atomic orbitals centred on the atom. The global density of states is equal to the sum of the local densities of states over all atoms and the electronic binding energy for each atom equals the integral of $d_i(E)E$:

$$E_i^{\text{el}} = \int d_i(E) E dE \quad (4.115)$$

Thus, if we knew the second moment of the local density of states we should be able to determine the atomic binding energy via the square root relationship. However, as quantum

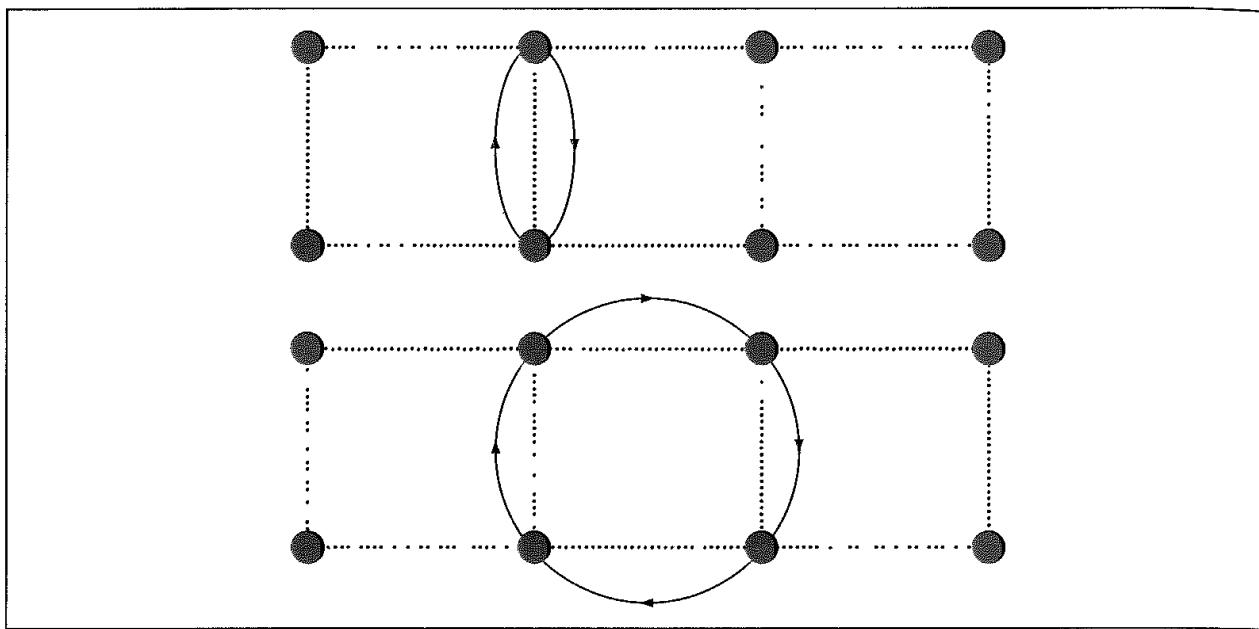


Fig 4.52. Calculating paths using the moments theorem. Illustrated are paths of lengths 2 and 4

mechanics is the only way we currently know of to determine the density of states, this might seem rather self-defeating. This is the role of the *moments theorem*, which relates the bonding topology to the moments of the local density of states without requiring an explicit calculation of the electronic energy levels.

The moments theorem states that the m th moment of the local density of states on an atom i is determined by the sum of all paths of length m over neighbouring atoms that start and end at i . For the second moment these paths involve just two ‘hops’, from the atom in question to a neighbour and back again (Figure 4.52). For the higher moments, the number of possible paths increases dramatically and becomes a challenging calculation. However, for the second moment the number of paths of length 2 is simply equal to the number of nearest neighbours, Z . Consequently, the local electronic binding energy for each atom is approximately equal to the square root of the number of neighbours. This is the *second-moment approximation*:

$$E_i^{\text{el}} \propto \sqrt{Z_i} \quad (4.116)$$

As an aside, we can easily show how this satisfies the ratio E_v/E_c (property 2, page 240) The energy E_v associated with Z atoms having their coordination reduced from Z to $Z - 1$ will be $Z[\sqrt{Z} - \sqrt{Z - 1}]$. The cohesive energy E_c is proportional to \sqrt{Z} For typical values of Z this gives E_v/E_c as approximately $\frac{1}{2}$.

In the Finnis–Sinclair potential a pairwise contribution is added to the many-body term to give the following form:

$$\mathcal{V} = \sum_{i=1}^N \sum_{j=i+1}^N P(r_{ij}) + \sum_{i=1}^N A\sqrt{\rho_i} \quad (4.117)$$

$P(r_{ij})$ is the pairwise potential, which, depending upon the model, can be considered to include electrostatic and repulsive contributions. The second term is a function of the electron density, ρ_i , and varies with the square root, in keeping with the second-moment approximation. The electron density for an atom includes contributions from the neighbouring atoms as follows:

$$\rho_i = \sum_{j=1, j \neq i}^N \phi_{ij}(r_{ij}) \quad (4.118)$$

$\phi_{ij}(r_{ij})$ is a short-range, decreasing function of the distance between the two atoms i and j . In the original Finnis-Sinclair model the function $\phi_{ij}(r_{ij})$ was written as a parabolic function of the interatomic distance, $(r_{ij} - r_c)^2$, where r_c is a cutoff distance chosen to lie between the second and third neighbouring shells. ϕ_{ij} is zero beyond this cutoff distance. The pairwise potential was expressed as a quartic polynomial up to some cutoff and zero beyond.

The Finnis-Sinclair potential can be written in a more general form by replacing the number of neighbouring atoms by an exponential function of the distance between atoms. This is necessary because the number of neighbours is not always straightforward to define, especially in disordered systems and near defects. An exponential function also reflects the fact that electron densities decay exponentially from the nucleus. Moreover, the pairwise potential can also be written as an exponential function of distance to give the following general equation:

$$\mathcal{V} = \sum_{i=1}^N \left\{ \sum_{j=1, j \neq i}^N A e^{-\alpha r_{ij}} - B \left[\sum_{j=1, j \neq i}^N e^{-\beta r_{ij}} \right]^{1/2} \right\} \quad (4.119)$$

Sutton and Chen extended the potential to longer range to enable the study of certain problems such as the interactions between clusters of atoms [Sutton and Chen 1990]. Their objective was to combine the superior Finnis-Sinclair description of short-range interactions with a van der Waals tail to model the long-range interactions. The form of the Sutton-Chen potential is:

$$\mathcal{V} = \varepsilon \left\{ \sum_{i=1}^N \sum_{j=i+1}^N \left(\frac{a}{r_{ij}} \right)^n - c \sum_{i=1}^N \left[\sum_{j=1, j \neq i}^N \left(\frac{a}{r_{ij}} \right)^m \right]^{1/2} \right\} \quad (4.120)$$

In this equation, ε and a are parameters with dimensions of energy and length respectively, c is a dimensionless (positive) parameter, and m and n are integers such that n is greater than m . The use of power-law relationships in the Sutton-Chen potential has a number of useful consequences, analogous to the scaling properties of the Lennard-Jones potential. For example, for a given crystal structure (e.g. hexagonal close-packed, face-centred cubic, body-centred cubic, etc.) the value of c is fixed. Moreover, if two metals are described by the same values of m and n then the results for one system may be converted directly to the other by rescaling the energy and length parameters ε and a . Typical values for m are between 6 and 8 and for n between 9 and 12.

The embedded-atom method [Daw and Baskes 1984] is an empirical embodiment of a simplified quantum mechanical model for bonding in solids called *effective medium*

theory. The key feature of effective medium theory is the replacement of the complex environment around each atom by a simplified model known as jellium. The jellium environment corresponds to a homogeneous electron gas with a positive background. Each atom is considered to be surrounded by a sphere with a radius such that the electronic charge within each sphere due to the background jellium is equal and opposite to the charge on the atom. In the embedded-atom method the background electron density is replaced by a sum of electron densities from the neighbouring atoms. The many-body term is known as an *embedding function*; this gives the energy of each atom as a function of the electron density, ρ_i . In the embedded-atom method the electron density ρ_i equals the sum of the electron densities ϕ_{ij} from neighbouring atoms (Equation (4.118)). In the Daw and Baskes model a Coulomb potential was used for the pairwise potential but with an effective charge $Z(r)$ that decreases gradually with internuclear distance. The embedding function was represented with a cubic spline equation that has a single minimum and goes to zero at vanishing density. The densities were obtained from quantum mechanical calculations.

Both the Finnis-Sinclair and the embedded-atom potentials (together with others that we have not considered here) can be represented using a very similar functional form. However, it is important to realise that they differ in the way that they connect to the first-principles, quantum mechanical model of bonding. They also differ in the procedures used to parametrise the models, so that different parametrisations may be reported for the same material.

The construction of empirical potentials for semiconductors is considered to be an even greater challenge than for metals. In our earlier discussion of the use of density functional methods to determine the electronic structure of the group 14 elements carbon, silicon and germanium we referred to the fact that, whilst the most stable form of silicon is the diamond structure, as pressure is applied so new structures can be obtained. That such a variety of structures can be achieved indicates that they are rather close in energy. Another interesting property of silicon is that in the liquid form it is a metal and the liquid is more dense than the solid. Two of the potentials that have been applied to these systems are the Stillinger-Weber and the Tersoff potentials. The Stillinger-Weber potential [Stillinger and Weber 1985] uses a two-body and three-body term:

$$\mathcal{V} = \sum_{i=1}^N \sum_{j=i+1}^N f_2(r_{ij}) + \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k=j+1}^N [h(r_{ij}, r_{ik}, \theta_{jik}) + h(r_{ji}, r_{jk}, \theta_{ijk}) + h(r_{ki}, r_{kj}, \theta_{ikj})] \quad (4.121)$$

$$f_2(r_{ij}) = A(Br_{ij}^{-p} - r_{ij}^{-q}) \exp[(r_i - a)^{-1}] \quad (4.122)$$

$$h(r_{ij}, r_{ik}, \theta_{jik}) = \lambda \exp[\gamma(r_{ij} - a)^{-1} + \gamma(r_{ik} - a)^{-1}] (\cos \theta_{jik} + \frac{1}{3})^2 \quad (4.123)$$

These equations all use distances and energies in reduced units and the functional form is designed to go to zero without discontinuities at the cutoff distance $r = a$. There are seven parameters ($A, B, p, q, a, \lambda, \gamma$), which were determined by a search procedure, with care being taken to ensure that the diamond structure was the most stable periodic arrangement and that the melting point and liquid structure (as determined by molecular dynamics simulations) were in reasonable agreement with experiment. The three-body term is

designed to favour the tetrahedral geometry found in the diamond structure, which is why it works reasonably well for this form of crystalline silicon. However, it does not perform so well for the other solid forms, which have a different atomic geometry, or for other properties such as the liquid structure.

The Tersoff potential [Tersoff 1988] is based on a model known as the *empirical bond-order potential*. This potential can be written in a form very similar to the Finnis–Sinclair potential:

$$\mathcal{V} = \sum_{i=1}^N \left\{ \sum_{j=1, j \neq i}^N A e^{\alpha r_{ij}} - b_{ij} B e^{-\beta r_{ij}} \right\} \quad (4.124)$$

The key term is b_{ij} , which is the bond order between the atoms i and j . This parameter depends upon the number of bonds to the atom i ; the strength of the ‘bond’ between i and j decreases as the number of bonds to the atom i increases. The original bond-order potential [Abell 1985] is mathematically equivalent to the Finnis–Sinclair model if the bond order b_{ij} is given by:

$$b_{ij} = \left(1 + \sum_{k=1; k \neq i, k \neq j}^N e^{-\beta(r_{ik} - r_{ij})} \right)^{-1/2} \quad (4.125)$$

It can be readily confirmed that b_{ij} decreases as the number of bonds N increases and/or their length (r_{ik}) decreases. This relationship between the bond strength and the number of neighbours provides a useful way to rationalise the structure of solids. Thus the high coordination of metals suggests that it is more effective for them to form more bonds, even though each individual bond is weakened as a consequence. Materials such as silicon achieve the balance for an intermediate number of neighbours and molecular solids have the smallest atomic coordination numbers.

The Tersoff potential was designed specifically for the group 14 elements and extends the basic empirical bond-order model by including an angular term. The interaction energy between two atoms i and j using this potential is:

$$\nu_{ij} = f_C(r_{ij})[A e^{-\lambda_1 r_{ij}} - b_{ij} B e^{-\lambda_2 r_{ij}}]$$

where

$$b_{ij} = (1 + \beta^n \zeta_{ij}^n)^{-1/2n}; \quad \zeta_{ij} = \sum_{k \neq i, j} f_C(r_{ik}) g(\theta_{ijk}) \exp[\lambda_3^3 (r_{ij} - r_{ik})^3] \quad (4.126)$$

$$g(\theta) = 1 + \frac{c^2}{d^2} - \frac{c^2}{[d^2 + (h - \cos \theta)^2]}$$

The function f_C is a smoothing function with the value 1 up to some distance r_{ij} (typically chosen to include just the first neighbour shell) and then smoothly tapers to zero at the cutoff distance. b_{ij} is the bond-order term, which incorporates an angular term dependent upon the bond angle θ_{ijk} . The Tersoff potential is more broadly applicable than the Stillinger–Weber potential, but does contain more parameters.

Appendix 4.1 The Interaction Between Two Drude Molecules

In the system comprising two Drude molecules (see Section 4.9.1), an additional term must be included in the Hamiltonian [Rigby *et al.* 1986]. This additional term arises from the interactions between the two dipoles. The instantaneous dipole of each molecule is $qz(t)$, where $z(t)$ is the separation of the charges. Thus, if we label the molecules 1 and 2, we can write the dipole-dipole interaction energy as:

$$\nu(\mu_1, \mu_2) = -\frac{2\mu_1\mu_2}{4\pi\varepsilon_0 r^3} = -\frac{2z_1 z_2 q^2}{4\pi\varepsilon_0 r^3} \quad (4.127)$$

r is the separation of the two molecules. The Schrödinger equation for this system is thus:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial z_1^2} - \frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial z_2^2} + \left[\frac{1}{2} k z_1^2 + \frac{1}{2} k z_2^2 - \frac{2z_1 z_2 q^2}{4\pi\varepsilon_0 r^3} \right] \psi = E\psi \quad (4.128)$$

This equation can be solved by making the following substitutions:

$$a_1 = \frac{z_1 + z_2}{\sqrt{2}}; \quad a_2 = \frac{z_1 - z_2}{\sqrt{2}}; \quad k_1 = k - \frac{2q^2}{4\pi\varepsilon_0 r^3}; \quad k_2 = k + \frac{2q^2}{4\pi\varepsilon_0 r^3} \quad (4.129)$$

These reduce Equation (4.128) to

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial a_1^2} - \frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial a_2^2} + [\frac{1}{2} k_1 a_1^2 + \frac{1}{2} k_2 a_2^2] \psi = E\psi \quad (4.130)$$

This is the Schrödinger equation for two independent (i.e. non-interacting) oscillators with frequencies given as follows:

$$\omega_1 = \omega \sqrt{1 - \frac{2q^2}{4\pi\varepsilon_0 r^3 k}}; \quad \omega_2 = \omega \sqrt{1 + \frac{2q^2}{4\pi\varepsilon_0 r^3 k}} \quad (4.131)$$

$\omega/2\pi$ is the frequency of an isolated Drude molecule. The ground state energy of the system is therefore just the sum of the zero-point energies of the two oscillators: $E_0 = \frac{1}{2}\hbar(\omega_1 + \omega_2)$

If we now substitute for ω_1 and ω_2 and expand the square roots using the binomial theorem, then we obtain the following:

$$E_0(r) = \hbar\omega - \frac{q^4 \hbar \omega}{2(4\pi\varepsilon_0)^2 r^6 k^2} - \dots \quad (4.132)$$

The interaction energy of the two oscillators is the difference between this zero-point energy and the energy of the system when the oscillators are infinitely separated and so:

$$\nu(r) = -\frac{q^4 \hbar \omega}{2(4\pi\varepsilon_0)^2 r^6 k^2} \quad (4.133)$$

The force constant, k , is related to the polarisability of the molecule, α as follows. Suppose a single Drude molecule is exposed to an external electric field \mathbf{E} . In the electric field, a force $q\mathbf{E}$ acts on each charge (in opposite directions as the charges are of opposite sign). This force causes the charges to separate and equilibrium is reached when the restoring force due to the stretching of the bond (kz) is equal to the electrostatic force: $qE = kz$. This separation

of the charges is equivalent to a static dipole given by $\mu_{ind} = qz = q^2E/k$. However, the induced dipole is also related to the polarisability by $\mu_{ind} = \alpha E$. Thus the polarisability can be written in terms of the force constant k : $\alpha = q^2/k$. With this substitution the result for the Drude model in two dimensions is:

$$\nu(r) = -\frac{\alpha^4 \hbar \omega}{2(4\pi\epsilon_0)^2 r^6} \quad (4.134)$$

In three dimensions the equivalent result is:

$$\nu(r) = -\frac{3\alpha^4 \hbar \omega}{4(4\pi\epsilon_0)^2 r^6} \quad (4.135)$$

Further Reading

- Bowen J P and N L Allinger 1991. Molecular Mechanics. The Art and Science of Parameterisation. In Lipkowitz K B and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 2. New York, VCH Publishers, pp 81-97
- Brenner D W, O A Shendreova and D A Areshkin 1998 Quantum-Based Analytic Interatomic Forces and Materials Simulation. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 12 New York, VCH Publishers, pp. 207-239.
- Burkert U and N L Allinger 1982. *Molecular Mechanics*. ACS Monograph 177. Washington D.C., American Chemical Society.
- Dykstra C E 1993. Electrostatic Interaction Potentials in Molecular Force Fields *Chemical Reviews* **93**:2339-2353
- Landis C R, D M Root and T Cleveland 1995 Molecular Mechanics Force Fields for Modeling Inorganic and Organometallic Compounds. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 6 New York, VCH Publishers, pp. 73-148
- Niketic S R and K Rasmussen 1977. *The Consistent Force Field. A Documentation*. Berlin, Springer-Verlag
- Price S L 2000 Towards More Accurate Model Intermolecular Potentials for Organic Molecules. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 14 New York, VCH Publishers, pp 225-289
- Rigby M, E B Smith, W A Wakeham and G C Maitland 1981 *Intermolecular Forces: Their Origin and Determination*. Oxford, Clarendon Press.
- Rigby M, E B Smith, W A Wakeham and G C Maitland 1986. *The Forces Between Molecules* Oxford, Clarendon Press.
- Van der Graaf B, S L Njo and K S Smirnov 2000. Introduction to Zeolite Modeling In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 14 New York, VCH Publishers, pp. 137-223
- Williams D E 1991. Net Atomic Charge and Multipole Models for the *Ab Initio* Molecular Electric Potential In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 2 New York, VCH Publishers, pp. 219-271.

References

- Abell G C 1985. Empirical Chemical Pseudopotential Theory of Molecular and Metallic Bonding *Physical Review B* **31**:6184-6196.

- Allan, N L and W C Mackrodt 1994 Density Functional Theory and Interionic Potentials *Philosophical Magazine* **B69**:871–878
- Allinger N L 1977. Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V₁ and V₂ Torsional Terms *Journal of the American Chemical Society* **99**:8127–8134.
- Allinger N L, K Chen and J-H Lii 1996a. An Improved Force Field (MM4) for Saturated Hydrocarbons *Journal of Computational Chemistry* **17**:642–668
- Allinger N L, K Chen, J A Katzenelenbogen, S R Wilson and G M Anstead 1996b. Hyperconjugative Effects on Carbon–Carbon Bond Lengths in Molecular Mechanics (MM4) *Journal of Computational Chemistry* **17**:747–755.
- Allinger N L, F Li and L Yan 1990a Molecular Mechanics The MM3 Force Field for Alkenes. *Journal of Computational Chemistry* **11**:848–867.
- Allinger N L, F Li, L Yan and J C Tai 1990b. Molecular Mechanics (MM3) Calculations on Conjugated Hydrocarbons. *Journal of Computational Chemistry* **11**:868–895
- Allinger N L and J T Sprague 1973 Calculation of the Structures of Hydrocarbons Containing Delocalised Electronic Systems by the Molecular Mechanics Method. *Journal of the American Chemical Society* **95** 3893–3907
- Allinger N L, Y H Yuh and J-J Lii 1989. Molecular Mechanics The MM3 Force Field for Hydrocarbons I. *Journal of the American Chemical Society* **111**:8551–9556
- Allured V S, C M Kelly and C R Landis 1991 SHAPES Empirical Force-Field – New Treatment of Angular Potentials and Its Application to Square-Planar Transition-Metal Complexes. *Journal of the American Chemical Society* **113**:1–12
- Barker J A, R A Fisher and R O Watts 1971. Liquid Argon: Monte Carlo and Molecular Dynamics Calculations *Molecular Physics* **21**:657–673
- Barnes P, J L Finney, J D Nicholas and J E Quinn 1979 Cooperative Effects in Simulated Water *Nature* **282**:459–464.
- Bayly C I, P Cieplak, W D Cornell and P A Kollman 1993. A Well-Behaved Electrostatic Potential Based Method for Deriving Atomic Charges – The RESP Model. *Journal of Physical Chemistry* **97**:10269–10280.
- Berendsen H C, J P M Postma, W F van Gunsteren and J Hermans 1981 Interaction Models for Water in Relation to Protein Hydration. In Pullman B (Editor) *Intermolecular Forces* Dordrecht, Reidel, pp 331–342.
- Berendsen H J C, J R Grigera and T P Straatsma 1987. The Missing Term in Effective Pair Potentials *Journal of Physical Chemistry* **91**:6269–6271.
- Bernal J D and R H Fowler 1933. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions *Journal of Chemical Physics* **1**:515–548.
- Bezler B H, K M Merz Jr and P A Kollman 1990. Atomic Charges Derived from Semi-Empirical Methods *Journal of Computational Chemistry* **11**:431–439
- Born M 1920. Volumen und Hydratationswärme der Ionen *Zeitschrift für Physik* **1**:45–48.
- Breneman C M and K B Wiberg 1990 Determining Atom-Centred Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis *Journal of Computational Chemistry* **11**:361–373.
- Buckingham A D 1959 Molecular Quadrupole Moments. *Quarterly Reviews of the Chemical Society* **13**:183–214.
- Churlan L E and M M Franci 1987. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study *Journal of Computational Chemistry* **8**:894–905.
- Claessens M, M Ferrario and J-P Ryckaert 1983 The Structure of Liquid Benzene. *Molecular Physics* **50**:217–227.
- Cleveland T and C R Landis 1996. Valence Bond Concepts Applied to the Molecular Mechanics Description of Molecular Shapes. 2 Applications to Hypervalent Molecules of the P-Block *Journal of the American Chemical Society* **118** 6020–6030.

- Corey E J and J C Bailar Jr 1959 The Stereochemistry of Complex Inorganic Compounds XXII Stereospecific Effects in Complex Ions. *Journal of the American Chemical Society* **81**:2620–2629
- Cornell W D, P Cieplak, C I Bayly, I R Gould, K M Merz Jr, D M Ferguson, D C Spellmeyer, T Fox, J W Caldwell and P A Kollman 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules *Journal of the American Chemical Society* **117**:5179–5197
- Corongiu G 1992. Molecular Dynamics Simulation for Liquid Water Using a Polarisable and Flexible Potential. *International Journal of Quantum Chemistry* **42** 1209–1235.
- Cox S R and D E Williams 1981. Representation of the Molecular Electrostatic Potential by a New Atomic Charge Model. *Journal of Computational Chemistry* **2**:304–323
- Dang L X, J E Rice, J Caldwell and P A Kollman 1991. Ion Solvation in Polarisable Water: Molecular Dynamics Simulations. *Journal of the American Chemical Society* **113** 2481–2486
- Daw M S and M I Baskes 1984. Embedded-atom Method: Derivation and Application to Impurities, Surfaces, and Other Defects in Metals *Physical Review* **B29**:6443–6453.
- Dick B G and A W Overhauser 1958 Theory of the Dielectric Constants of Alkali Halide Crystals *Physical Review* **112** 90–103.
- Dinur U and A T Hagler 1991. New Approaches to Empirical Force Fields. In K B Lipkowitz and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 2. New York, VCH Publishers, pp. 99–164.
- Dinur U and A T Hagler 1995. Geometry-Dependent Atomic Charges Methodology and Application to Alkanes, Aldehydes, Ketones and Amides *Journal of Computational Chemistry* **16**:154–170.
- Ferenczy G G, C A Reynolds and W G Richards 1990 Semi-Empirical AM1 Electrostatic Potentials and AM1 Electrostatic Potential Derived Charges – A Comparison with *Ab Initio* Values. *Journal of Computational Chemistry* **11**:159–169.
- Ferguson D M 1995 Parameterisation and Evaluation of a Flexible Water Model. *Journal of Computational Chemistry* **16**.501–511.
- Finnis M W and J E Sinclair 1984. A Simple Empirical *N*-body Potential for Transition Metals. *Philosophical Magazine* **A50** 45–55.
- Fowler P W and A D Buckingham 1991 Central or Distributed Multipole Moments? Electrostatic Models of Aromatic Dimers *Chemical Physics Letters* **176**:11–18.
- Gale J D, C R A Catlow and W C Mackrodt 1992. Periodic *Ab Initio* Determination of Interatomic Potentials for Alumina. *Modelling and Simulation in Materials Science and Engineering* **1** 73–81
- Gasteiger J and M Marsili 1980 Iterative Partial Equalization of Orbital Electronegativity – Rapid Access to Atomic Charges. *Tetrahedron* **36**.3219–3288.
- Gay J G and B J Berne 1981. Modification of the Overlap Potential to Mimic a Linear Site–Site Potential. *Journal of Chemical Physics* **74**:3316–3319.
- Goodford P J 1985 A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *Journal of Medicinal Chemistry* **28**.849–857
- Hagler A T, E Huler and S Lifson 1977. Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals *Journal of the American Chemical Society* **96** 5319–5327.
- Hagler A T and S Lifson 1974 Energy Functions for Peptides and Proteins II. The Amide Hydrogen Bond and Calculation of Amide Crystal Properties. *Journal of the American Chemical Society* **96**:5327–5335.
- Halgren T A 1992 Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields. Potential Form, Combination Rules, and vdW Parameters. *Journal of the American Chemical Society* **114**:7827–7843.
- Halgren T A 1996a. Merck Molecular Force Field I. Basis, Form, Scope, Parameterisation and Performance of MMFF94 *Journal of Computational Chemistry* **17**:490–519.
- Halgren T A 1996b. Merck Molecular Force Field II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *Journal of Computational Chemistry* **17**:520–552

- Hill T L 1948. Steric Effects I. Van der Waals Potential Energy Curves *Journal of Chemical Physics* **16**:399–404.
- Hunter C A 1993. Sequence-dependent DNA structure The role of base stacking interactions. *Journal of Molecular Biology* **230**:1024–1054
- Hunter C A and J K M Saunders 1990. The Nature of π - π Interactions *The Journal of the American Chemical Society* **112**:5525–5534.
- Hwang M J, T P Stockfisch and A T Hagler 1994. Derivation of Class II Force Fields 2 Derivation and Characterisation of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules. *Journal of the American Chemical Society* **116**:2515–2525.
- Jorgensen W L, J Chandrasekhar, J D Madura, R W Impey and M L Klein 1983 Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**:926–935.
- Jorgensen W L and J Pranata 1990 Importance of Secondary Interactions in Triply Hydrogen Bonded Complexes: Guanine–Cytosine vs Uracil–2,6-Diaminopyridine. *Journal of the American Chemical Society* **112**:2008–2010.
- Jorgensen W L and J Tirado-Rives 1988 The OPLS Potential Functions for Proteins – Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* **110**:1666–1671.
- Landis C R, T Cleveland and T K Firman 1995. Making Sense of the Shapes of Simple Metal Hydrides *Journal of the American Chemical Society* **117**:1859–1860
- Landis C R, T K Firman, D M Root and T Cleveland 1998. A Valence Bond Perspective on the Molecular Shapes of Simple Metal Alkyls and Hydrides *Journal of the American Chemical Society* **120**:1842–1854
- Lifson S and A Warshel 1968. Consistent Force Field for Calculations of Conformations, Vibrational Spectra and Enthalpies of Cycloalkane and *n*-Alkane Molecules *Journal of Chemical Physics* **49**:5116–5129.
- Lii J-H and N L Allinger 1989. Molecular Mechanics. The MM3 Force Field for Hydrocarbons 2 Vibrational Frequencies and Thermodynamics *Journal of the American Chemical Society* **111**:8566–8582
- London F 1930 Zur Theorie und Systematik der Molekularkräfte *Zeitschrift für Physik* **63**:245–279
- Luckhurst G R, R A Stephens and R W Phippen 1990 Computer Simulation Studies of Anisotropic Systems XIX Mesophases Formed by the Gay–Berne Model Mesogen *Liquid Crystals* **8**:451–464
- Luque F J, F Ilas and M Orozco 1990 Comparative Study of the Molecular Electrostatic Potential Obtained from Different Wavefunctions – Reliability of the Semi-Empirical MNDO Wavefunction *Journal of Computational Chemistry* **11**:416–430.
- Lybrand T P and P A Kollman 1985 Water-Water and Water-Ion Potential Functions Including Terms for Many Body Effects. *Journal of Chemical Physics* **83**:2923–2933
- Maple J R, U Dinur and A T Hagler 1988. Derivation of Force Fields for Molecular Mechanics and Molecular Dynamics from *Ab Initio* Energy Surfaces *Proceedings of the National Academy of Sciences USA* **85**:5350–5354
- Nevins N, K Chen and N L Allinger 1996a. Molecular Mechanics (MM4) Calculations on Alkenes. *Journal of Computational Chemistry* **17**:669–694.
- Nevins N, K Chen and N L Allinger 1996b. Molecular Mechanics (MM4) Calculations on Conjugated Hydrocarbons. *Journal of Computational Chemistry* **17**:695–729.
- Nevins N, K Chen and N L Allinger 1996c. Molecular Mechanics (MM4) Vibrational Frequency Calculations for Alkenes and Conjugated Hydrocarbons. *Journal of Computational Chemistry* **17**:730–746
- Nicholas J B, A J Hopfinger, F R Trouw and L E Iton 1991 Molecular Modelling of Zeolite Structure. 2. Structure and Dynamics of Silica Sodalite and Silicate Force Field. *The Journal of the American Chemical Society* **113**:4792–4800.
- Niesar U, G Corongiu, E Clementi, G R Keller and D K Bhattacharya 1990. Molecular Dynamics Simulations of Liquid Water Using the NCC *Ab Initio* Potential *Journal of Physical Chemistry* **94**:7949–7956.

- Niketic S R and K Rasmussen 1977 *The Consistent Force Field: A Documentation*. Berlin, Springer-Verlag
- Packer M J, M P Dauncey and C A Hunter 2000 Sequence-dependent DNA Structure Dinucleotide Conformational Maps *Journal of Molecular Biology* **295**:71–83.
- Pranata J and W L Jorgensen 1991. Computational Studies on FK506: Computational Search and Molecular Dynamics Simulations in Water. *Journal of the American Chemical Society* **113**:9483–9493.
- Price S L, R J Harrison and M F Guest 1989. An *Ab Initio* Distributed Multipole Study of the Electrostatic Potential Around an Undecapeptide Cyclosporin Derivative and a Comparison with Point Charge Electrostatic Models *Journal of Computational Chemistry* **10**:552–567.
- Rappé A K, C J Casewit, K S Colwell, W A Goddard III and W M Skiff 1992 UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations *Journal of the American Chemical Society* **114** 10024–10035.
- Rappé A K, K S Colwell and C J Casewit 1993 Application of a Universal Force Field to Metal Complexes. *Inorganic Chemistry* **32**:3438–3450.
- Rappé A K and W A Goddard III 1991. Charge Equilibration for Molecular Dynamics Simulations. *Journal of Physical Chemistry* **95**:3358–3363
- Reynolds C A, J W Essex and W G Richards 1992. Atomic Charges for Variable Molecular Conformations *Journal of the American Chemical Society* **114**:9075–9079.
- Rick S W and B J Berne 1996. Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. *Journal of the American Chemical Society* **118**:672–679.
- Rick S W, S J Stuart and B J Berne 1994. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *Journal of Chemical Physics* **101**:6141–6156.
- Rigby M, E B Smith, W A Wakeham and G C Maitland 1986 *The Forces Between Molecules*. Oxford, Clarendon Press.
- Rodger P M, A J Stone and D J Tildesley 1988 The Intermolecular Potential of Chlorine. A Three Phase Study. *Molecular Physics* **63**:173–188
- Singh U C and P A Kollman 1984. An Approach to Computing Electrostatic Charges for Molecules. *Journal of Computational Chemistry* **5**:129–145
- Smith P E and B M Pettitt 1994. Modelling Solvent in Biomolecular Systems. *Journal of Physical Chemistry* **98**:9700–9711.
- Sprague J T, J C Tai, Y Yuh and N L Allinger 1987 The MMP2 Calculational Method *Journal of Computational Chemistry* **8**:581–603
- Sprik M and M L Klein 1988. A Polarisable Model for Water Using Distributed Charge Sites. *Journal of Chemical Physics* **89**:7556–7560
- Stillinger F H and A Rahman 1974 Improved Simulation of Liquid Water by Molecular Dynamics. *Journal of Chemical Physics* **60**:1545–1557.
- Stillinger F H and T A Weber 1985. Computer Simulation of Local Order in Condensed Phases of Silicon. *Physical Review B* **31**:5262–5271
- Stone A J 1981. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution *Chemical Physics Letters* **83**:233–239
- Stone A J and M Alderton 1985 Distributed Multipole Analysis Methods and Applications *Molecular Physics* **56**:1047–1064.
- Stuart S J and B J Berne 1996. Effects of Polarisability on the Hydration of the Chloride Ion. *Journal of Physical Chemistry* **100**:11934–11943.
- Sutton A P and J Chen 1990. Long-range Finnis-Sinclair Potentials. *Philosophical Magazine Letters* **61**:139–146.
- Tersoff J 1988. New Empirical Approach for the Structure and Energy of Covalent Systems. *Physical Review B* **37**:6991–7000
- Toxvaerd S 1990. Molecular Dynamics Calculation of the Equation of State of Alkanes *Journal of Chemical Physics* **93**:4290–4295

- Vedani A 1988. YETI: An Interactive Molecular Mechanics Program for Small-Molecular Protein Complexes *Journal of Computational Chemistry* **9**:269–280.
- Vinter J G 1994. Extended Electron Distributions Applied to the Molecular Mechanics of Some Intermolecular Interactions *Journal of Computer-Aided Molecular Design* **8**:653–668.
- Warshel A and M Karplus 1972. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I Formulation and Parameterisation *Journal of the American Chemical Society* **94**.5612-5622.
- Warshel A and A Lappicirella 1981. Calculations for Ground- and Excited-State Potential Surfaces for Conjugated Heteroatomic Molecules. *Journal of the American Chemical Society* **103**.4664-4673
- Weiner S J, P A Kollman, D A Case, U C Singh, C Ghio, G Alagona, S Profeta and P Weiner 1984 A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *Journal of the American Chemical Society* **106**.765–784.
- Williams D E 1990. Alanyl Dipeptide Potential-Derived Net Atomic Charges and Bond Dipoles, and Their Variation with Molecular Conformation. *Biopolymers* **29**:1367–1386

Energy Minimisation and Related Methods for Exploring the Energy Surface

5.1 Introduction

For all except the very simplest systems the potential energy is a complicated, multi-dimensional function of the coordinates. For example, the energy of a conformation of ethane is a function of the 18 internal coordinates or 24 Cartesian coordinates that are required to completely specify the structure. As we discussed in Section 1.3, the way in which the energy varies with the coordinates is usually referred to as the *potential energy surface* (sometimes called the *hypersurface*). In the interests of brevity all references to ‘energy’ should be taken to mean ‘potential energy’ for the rest of this chapter, except where explicitly stated otherwise. For a system with N atoms the energy is thus a function of $3N - 6$ internal or $3N$ Cartesian coordinates. It is therefore impossible to visualise the entire energy surface except for some simple cases where the energy is a function of just one or two coordinates. For example, the van der Waals energy of two argon atoms (as might be modelled using the Lennard-Jones potential function) depends upon just one coordinate: the interatomic distance. Sometimes we may wish to visualise just a part of the energy surface. For example, suppose we take an extended conformation of pentane and rotate the two central carbon–carbon bonds so that the torsion angles vary from 0° to 360° , calculating the energy of each structure generated. The energy in this case is a function of just two variables and can be plotted as a contour diagram or as an isometric plot, as shown in Figure 5.1.

We will use the term ‘energy surface’ to refer not only to systems in which the bonding remains unchanged, as in these two examples, but also where bonds are broken and/or formed. Our discussion will be appropriate to both quantum mechanics and molecular mechanics, except where otherwise stated.

In molecular modelling we are especially interested in minimum points on the energy surface. Minimum energy arrangements of the atoms correspond to stable states of the system; any movement away from a minimum gives a configuration with a higher energy. There may be a very large number of minima on the energy surface. The minimum with the very lowest energy is known as the *global energy minimum*. To identify those geometries of

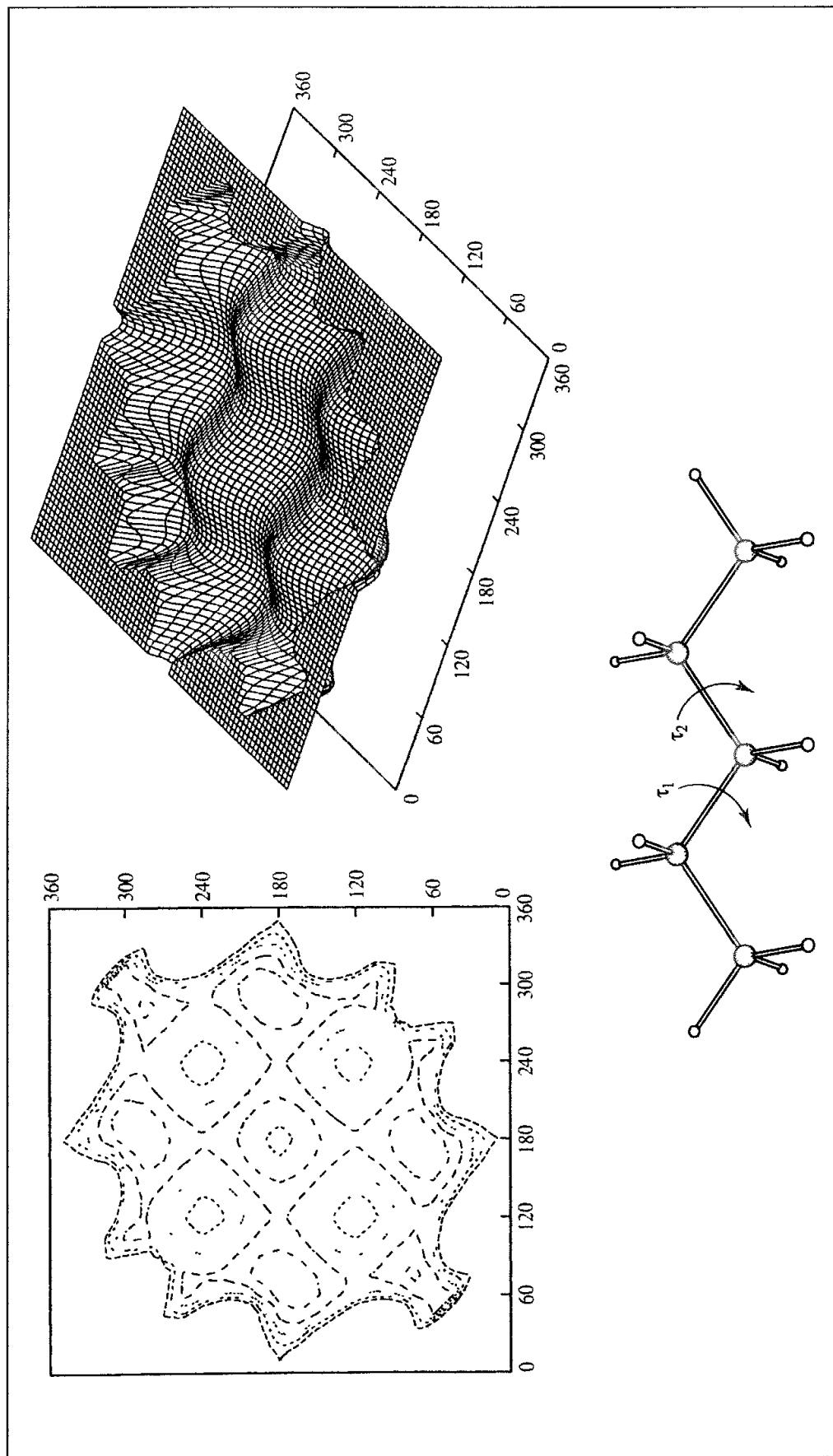


Fig. 5.1 Variation in the energy of pentane with the two torsion angles indicated and represented as a contour diagram and isometric plot. Only the lowest-energy regions are shown.

the system that correspond to minimum points on the energy surface we use a *minimisation algorithm*. There is a vast literature on such methods and so we will concentrate on those approaches that are most commonly used in molecular modelling. We may also be interested to know how the system changes from one minimum energy structure to another. For example, how do the relative positions of the atoms vary during a reaction? What structural changes occur as a molecule changes its conformation? The highest point on the pathway between two minima is of especial interest and is known as the *saddle point*, with the arrangement of the atoms being the *transition structure*. Both minima and saddle points are stationary points on the energy surface, where the first derivative of the energy function is zero with respect to all the coordinates.

A geographical analogy can be a helpful way to illustrate many of the concepts we shall encounter in this chapter. In this analogy minimum points correspond to the bottom of valleys. A minimum may be described as being in a 'long and narrow valley' or 'a flat and featureless plain'. Saddle points correspond to mountain passes. We refer to algorithms taking steps 'uphill' or 'downhill'.

5.1.1 Energy Minimisation: Statement of the Problem

The minimisation problem can be formally stated as follows: given a function f which depends on one or more independent variables x_1, x_2, \dots, x_i , find the values of those variables where f has a minimum value. At a minimum point the first derivative of the function with respect to each of the variables is zero and the second derivatives are all positive:

$$\frac{\partial f}{\partial x_i} = 0; \quad \frac{\partial^2 f}{\partial x_i^2} > 0 \quad (5.1)$$

The functions of most interest to us will be the quantum mechanics or molecular mechanics energy with the variables x_i being the Cartesian or the internal coordinates of the atoms. Molecular mechanics minimisations are nearly always performed in Cartesian coordinates, where the energy is a function of $3N$ variables; it is more common to use internal coordinates (as defined in the Z-matrix) with quantum mechanics. For analytical functions, the minimum of a function can be found using standard calculus methods. However, this is not generally possible for molecular systems due to the complicated way in which the energy varies with the coordinates. Rather, minima are located using numerical methods, which gradually change the coordinates to produce configurations with lower and lower energies until the minimum is reached. To illustrate how the various minimisation algorithms operate, we shall consider a simple function of two variables: $f(x, y) = x^2 + 2y^2$. This function is represented as a contour diagram in Figure 5.2. The function has one minimum point, located at the origin. In our examples we will attempt to locate the minimum from the point (9.0, 9.0). Although this is a function of just two variables for the purposes of illustration, all of the methods that we shall consider can be applied to functions of many more variables.

We can classify minimisation algorithms into two groups: those which use derivatives of the energy with respect to the coordinates and those which do not. Derivatives can be useful

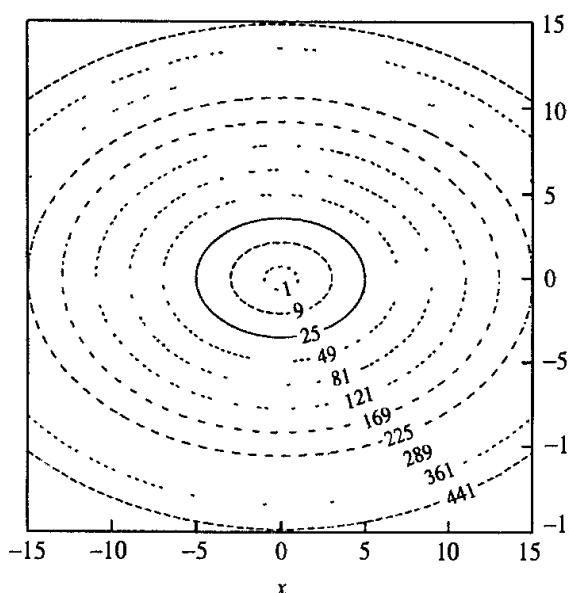


Fig. 5.2. The function $x^2 + 2y^2$.

because they provide information about the shape of the energy surface, and, if used properly, they can significantly enhance the efficiency with which the minimum is located. There are many factors that must be taken into account when choosing the most appropriate algorithm (or combination of algorithms) for a given problem; the ideal minimisation algorithm is the one that provides the answer as quickly as possible, using the least amount of memory. No single minimisation method has yet proved to be the best for all molecular modelling problems and so most software packages offer a choice of methods. In particular, a method that works well with quantum mechanics may not be the most suitable for use with molecular mechanics. This is partly because quantum mechanics is usually used to model systems with fewer atoms than molecular mechanics; some operations that are integral to certain minimisation procedures (such as matrix inversion) are trivial for small systems but formidable for systems containing thousands of atoms. Quantum mechanics and molecular mechanics also require different amounts of computational effort to calculate the energies and the derivatives of the various configurations. Thus an algorithm that takes many steps may be appropriate for molecular mechanics but inappropriate for quantum mechanics.

Most minimisation algorithms can only go downhill on the energy surface and so they can only locate the minimum that is nearest (in a downhill sense) to the starting point. Thus, Figure 5.3 shows a schematic energy surface and the minima that would be obtained starting from three points A, B and C. The minima can be considered to correspond to the locations where a ball rolling on the energy surface under the influence of gravity would come to rest. To locate more than one minimum or to locate the global energy minimum we therefore usually require a means of generating different starting points, each of which is then minimised. Some specialised minimisation methods can make uphill moves to seek out minima lower in energy than the nearest one, but no algorithm has yet proved capable of locating the

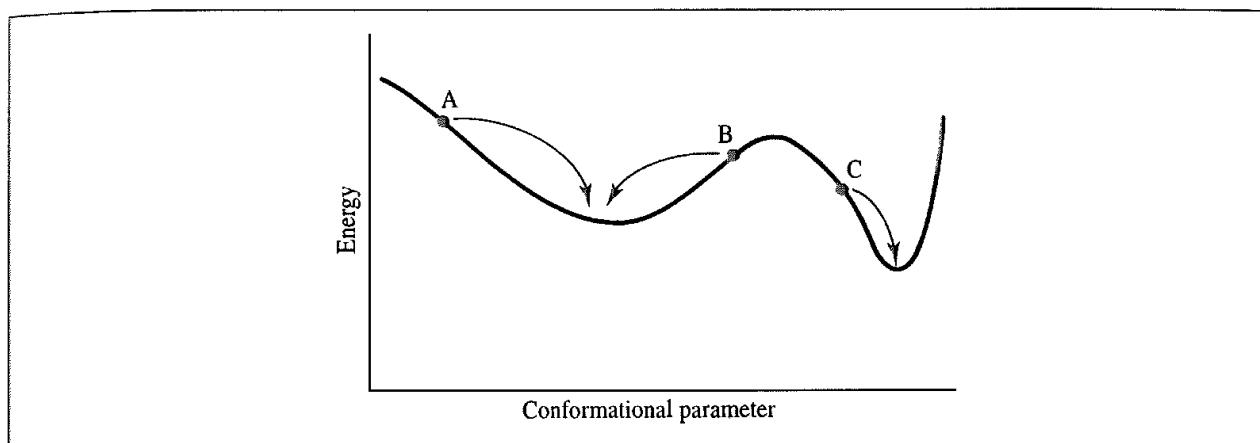


Fig 5.3: A schematic one-dimensional energy surface. Minimisation methods move downhill to the nearest minimum. The statistical weight of the narrow, deep minimum may be less than a broad minimum which is higher in energy.

global energy minimum from an arbitrary starting position. The shape of the energy surface may be important if one wishes to calculate the relative populations of the various minimum energy structures. For example, a deep and narrow minimum may be less highly populated than a broad minimum that is higher in energy as the vibrational energy levels will be more widely spaced in the deeper minimum and so less accessible. For this reason the global energy minimum may not be the most highly populated minimum. In any case, the 'active' structure (e.g. the biologically active conformation of a drug molecule) may not correspond to the global minimum, or to the most highly populated conformation, or even to a minimum energy structure at all.

The input to a minimisation program consists of a set of initial coordinates for the system. The initial coordinates may come from a variety of sources. They may be obtained from an experimental technique, such as X-ray crystallography or NMR. In other cases a theoretical method is employed, such as a conformational search algorithm. A combination of experimental and theoretical approaches may also be used. For example, to study the behaviour of a protein in water one may take an X-ray structure of the protein and immerse it in a solvent 'bath', where the coordinates of the solvent molecules have been obtained from a Monte Carlo or molecular dynamics simulation.

5.1.2 Derivatives

In order to use a derivative minimisation method it is obviously necessary to be able to calculate the derivatives of the energy with respect to the variables (i.e. the Cartesian or internal coordinates, as appropriate). Derivatives may be obtained either analytically or numerically. The use of analytical derivatives is preferable as they are exact, and because they can be calculated more quickly; if only numerical derivatives are available then it may be more effective to use a non-derivative minimisation algorithm. The problems of calculating analytical derivatives with quantum mechanics and molecular mechanics were discussed in Sections 3.4.3 and 4.16, respectively.

Nevertheless, under some circumstances it is necessary to use numerical derivatives. These can be calculated as follows. If one of the coordinates x_i is changed by a small change (δx_i) and the energy for the new arrangement is computed then the derivative $\partial E / \partial x_i$ is obtained by dividing the change in energy (δE) by the change in coordinate ($\delta E / \delta x_i$). This strictly gives the derivative at the mid-point between the two points x_i and $x_i + \delta x_i$. A more accurate value of the derivative at the point x_i may be obtained (at the cost of an additional energy calculation) by evaluating the energy at two points, $x_i + \delta x_i$ and $x_i - \delta x_i$. The derivative is then obtained by dividing the difference in the energies by $2\delta x_i$.

5.2 Non-derivative Minimisation Methods

5.2.1 The Simplex Method

A *simplex* is a geometrical figure with $M + 1$ interconnected vertices, where M is the dimensionality of the energy function. For a function of two variables the simplex is thus triangular in shape. A tetrahedral simplex is used for a function of three variables and so for an energy function of $3N$ Cartesian coordinates the simplex will have $3N + 1$ vertices; if internal coordinates are used then the simplex will have $3N - 5$ vertices. Each vertex corresponds to a specific set of coordinates for which an energy can be calculated. For our function $f(x, y) = x^2 + 2y^2$ the simplex method would use a triangular simplex.

The simplex algorithm locates a minimum by moving around on the potential energy surface in a fashion that has been likened to the motion of an amoeba. Three basic kinds of move are possible. The most common type of move is a reflection of the vertex with the highest value through the opposite face of the simplex, in an attempt to generate a new point that has a lower value. If this new point is lower in energy than any of the other points in the simplex then a ‘reflection and expansion’ move may be applied. When a ‘valley floor’ is reached then a reflection move will fail to produce a better point. Under such circumstances the simplex contracts along one dimension from the highest point. If this fails to reduce the energy then a third type of move is possible, in which the simplex contracts in all directions, pulling around the lowest point. These three moves are illustrated in Figure 5.4.

To implement the simplex algorithm it is first necessary to generate the vertices of the initial simplex. The initial configuration of the system corresponds to just one of these vertices. The remaining points can be obtained in a variety of ways, but one simple method is to add a constant increment to each coordinate in turn. The energy of the system is calculated at the new point, giving the function value for the relevant vertex.

The simplex method is most useful where the initial configuration of the system is very high in energy, because it rarely fails to find a better solution. However, it can be rather expensive in terms of computer time due to the large number of energy evaluations which are required (merely to generate the initial simplex requires $3N + 1$ energy evaluations). For this reason the simplex method is often used in combination with a different minimisation algorithm: a few steps of the simplex method are used to refine the initial structure and then a more efficient method can take over.

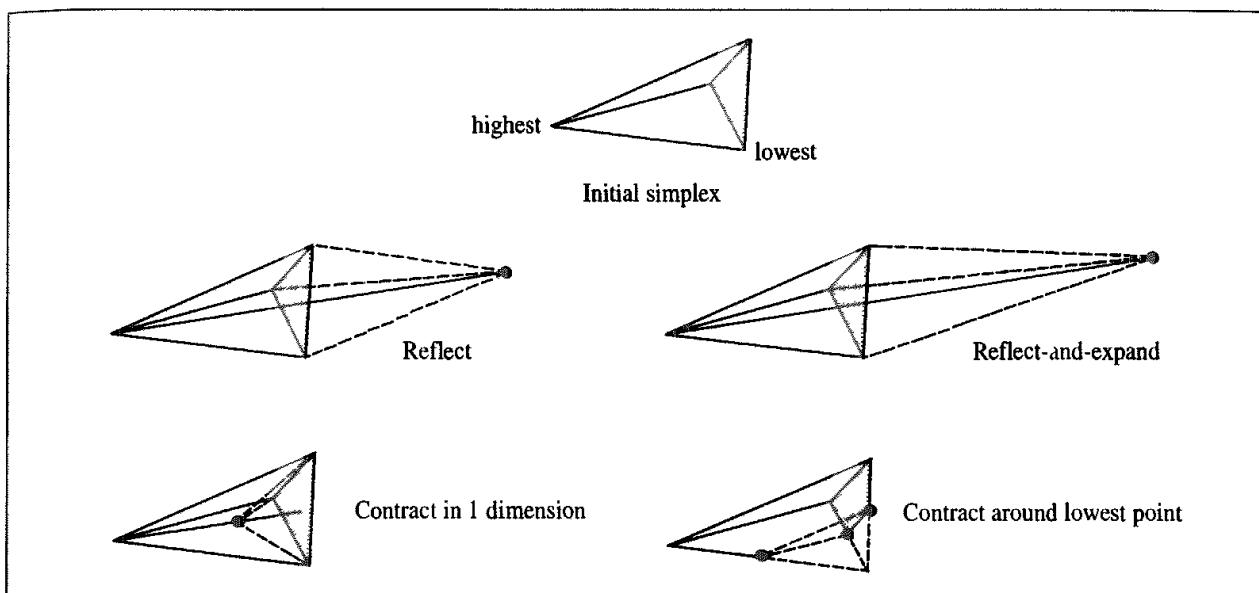


Fig 5.4: The three basic moves permitted to the simplex algorithm (reflection, and its close relation reflect-and-expand; contract in one dimension and contract around the lowest point) (Figure adapted from Press W H, B P Flannery, S A Teukolsky and W T Vetterling 1992. Numerical Recipes in Fortran. Cambridge, Cambridge University Press.)

Let us consider the application of the simplex method to our quadratic function, $f = x^2 + 2y^2$ (Figure 5.5). Suppose our initial simplex contains vertices located at the points (9, 9), (11, 9) and (9, 11), which have been generated by adding a constant factor 2 to each of the variables in turn. The values of the function at these points are 243, 283 and 323, respectively. The vertex with the highest function value is at (9, 11) and so in the first iteration this point is reflected through the opposite face of the triangle to generate a point with coordinates (11, 7) and a function value of 219 (we do not use the reflect-and-expand move in our

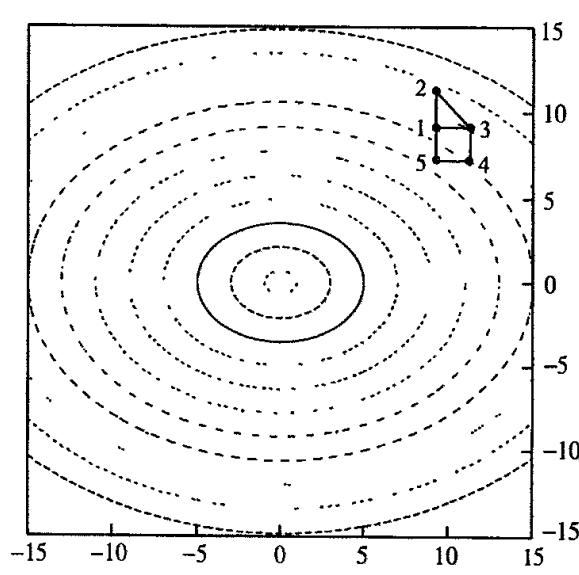


Fig 5.5: The first few steps of the simplex algorithm with the function $x^2 + 2y^2$. The initial simplex corresponds to the triangle 123. Point 2 has the largest value of the function and the next simplex is the triangle 134. The simplex for the third step is 145.

illustration). The highest vertex is now at (11, 9), which is reflected through the opposite face of the simplex to give the point (9, 7), where the function has a value of 179. In fact, for this admittedly artificial problem the simplex algorithm takes more than 30 steps to find a point where the function has a value less than 0.1.

Why does the simplex contain one more vertex than the number of degrees of freedom? The reason is that with fewer than $M + 1$ vertices the algorithm cannot explore the whole energy surface. Suppose we use only a two-vertex simplex to explore our quadratic energy surface. A simplex with just two vertices is a straight line. The only moves that would be possible in this case would be to other points that lie on this line; none of the energy surface away from the line would be explored. Similarly, if we have a function of three variables and restrict the simplex to a triangle then we will only be able to explore the region of space that lies in the same plane as the triangle, whereas the minimum may not lie in this plane.

5.2.2 The Sequential Univariate Method

The simplex method is rarely considered suitable for quantum mechanical calculations, due to the number of energy evaluations that must be performed. The sequential univariate method is a non-derivative method that is considered more appropriate in this case. This method systematically cycles through the coordinates in turn. For each coordinate, two new structures are generated by changing the current coordinate (i.e. $x_i + \delta x_i$ and $x_i + 2\delta x_i$). The energies of these two structures are calculated. A parabola is then fitted through the three points corresponding to the two distorted structures and the original structure. The minimum point in this quadratic function is determined and the coordinate is then changed to the position of the minimum. The procedure is illustrated in Figure 5.6. When the changes in all the coordinates are

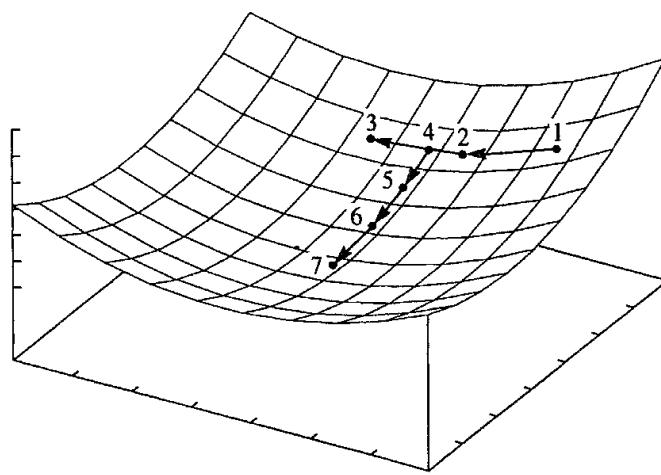


Fig. 5.6. The sequential univariate method. Starting at the point labelled 1 two steps are made along one of the coordinates to give points 2 and 3. A parabola is fitted to these three points and the minimum located (point 4). The same procedure is then repeated along the next coordinate (points 5, 6 and 7). (Figure adapted from Schlegel H B 1987. Optimization of Equilibrium Geometries and Transition Structures In Lawley K P (Editor) *Ab Initio Methods in Quantum Chemistry - I* New York, John Wiley, pp. 249–286)

sufficiently small then the minimum is deemed to have been reached, otherwise a new iteration is performed. The sequential invariate method usually requires fewer function evaluations than the simplex method but it can be slow to converge especially if there is strong coupling between two or more of the coordinates or when the energy surface is analogous to a long narrow valley.

5.3 Introduction to Derivative Minimisation Methods

Derivatives provide information that can be very useful in energy minimisation, and derivatives are used by most popular minimisation methods. The direction of the first derivative of the energy (the gradient) indicates where the minimum lies, and the magnitude of the gradient indicates the steepness of the local slope. The energy of the system can be lowered by moving each atom in response to the force acting on it; the force is equal to minus the gradient. Second derivatives indicate the curvature of the function, information that can be used to predict where the function will change direction (i.e. pass through a minimum or some other stationary point).

When discussing derivative methods it is useful to write the function as a Taylor series expansion about the point x_k :

$$\mathcal{V}(x) = \mathcal{V}(x_k) + (x - x_k)\mathcal{V}'(x_k) + (x - x_k)^2\mathcal{V}''(x_k)/2 + \dots \quad (5.2)$$

For a multidimensional function, the variable x is replaced by the vector \mathbf{x} and matrices are used for the various derivatives. Thus if the potential energy $\mathcal{V}(\mathbf{x})$ is a function of $3N$ Cartesian coordinates, the vector \mathbf{x} will have $3N$ components and \mathbf{x}_k corresponds to the current configuration of the system. $\mathcal{V}'(\mathbf{x}_k)$ is a $3N \times 1$ matrix (i.e. a vector), each element of which is the partial derivative of \mathcal{V} with respect to the appropriate coordinate, $\partial\mathcal{V}/\partial x_i$. We will also write the gradient at the point k as \mathbf{g}_k . Each element (i,j) of the matrix $\mathcal{V}''(\mathbf{x}_k)$ is the partial second derivative of the energy function with respect to the two coordinates x_i and x_j , $\partial^2\mathcal{V}/\partial x_i \partial x_j$. $\mathcal{V}''(\mathbf{x}_k)$ is thus of dimension $3N \times 3N$ and is known as the *Hessian* matrix or the *force constant* matrix. The Taylor series expansion can be written in the following form for the multidimensional case:

$$\mathcal{V}(\mathbf{x}) = \mathcal{V}(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)\mathcal{V}'(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)^T \cdot \mathcal{V}''(\mathbf{x}_k) \cdot (\mathbf{x} - \mathbf{x}_k)/2 + \dots \quad (5.3)$$

The energy functions used in molecular modelling are rarely quadratic and so the Taylor series expansion, Equation (5.3), can only be considered an approximation. There are two important consequences of this. The first consequence is that the performance of a given minimisation method will not be as good for a molecular mechanics or quantum mechanics energy surface as it is for a pure quadratic function. As we shall see, a second derivative method such as the Newton-Raphson algorithm can locate the minimum in a single step for a purely quadratic function, but several iterations are usually required for a typical molecular modelling energy function. The second consequence is that, far from the minimum, the harmonic approximation is a poor one and some of the less robust methods will fail, even though they may work very well close to a minimum, where the harmonic approximation is more valid. For this reason it is important to choose the minimisation

protocol with care, possibly using a robust (but perhaps inefficient) method at first, and then a less robust but more efficient method.

The derivative methods can be classified according to the highest-order derivative used. First-order methods use the first derivatives (i.e. the gradients) whereas second-order methods use both first and second derivatives. The simplex method can thus be considered a zeroth-order method as it does not use any derivatives.

5.4 First-order Minimisation Methods

Two first-order minimisation algorithms that are frequently used in molecular modelling are the method of *steepest descents* and the *conjugate gradient* method. These gradually change the coordinates of the atoms as they move the system closer and closer to the minimum point. The starting point for each iteration (k) is the molecular configuration obtained from the previous step, which is represented by the multidimensional vector \mathbf{x}_{k-1} . For the first iteration the starting point is the initial configuration of the system provided by the user, the vector \mathbf{x}_1 .

5.4.1 The Steepest Descents Method

The steepest descents method moves in the direction parallel to the net force, which in our geographical analogy corresponds to walking straight downhill. For $3N$ Cartesian coordinates this direction is most conveniently represented by a $3N$ -dimensional unit vector, \mathbf{s}_k . Thus:

$$\mathbf{s}_k = -\mathbf{g}_k / |\mathbf{g}_k| \quad (5.4)$$

Having defined the direction along which to move it is then necessary to decide how far to move along the gradient. Consider the two-dimensional energy surface of Figure 5.7. The gradient direction from the starting point is along the line indicated. If we imagine a cross-section through the surface along the line, the function will pass through a minimum and then increase, as shown in the figure. We can choose to locate the minimum point by performing a *line search* or we can take a step of arbitrary size along the direction of the force

5.4.2 Line Search in One Dimension

The purpose of a line search is to locate the minimum along a specified direction (i.e. along a line through the multidimensional space). The first stage of the line search is to *bracket* the minimum. This entails finding three points along the line such that the energy of the middle point is lower than the energy of the two outer points. If three such points can be found, then at least one minimum must lie between the two outer points. An iterative procedure can then be used to decrease the distance between the three points, gradually restricting the minimum to an even smaller region. This is conceptually an easy process but it may require a considerable number of function evaluations, making it computationally expensive. An alternative is to fit a function such as a quadratic to the three points. Differentiation of the

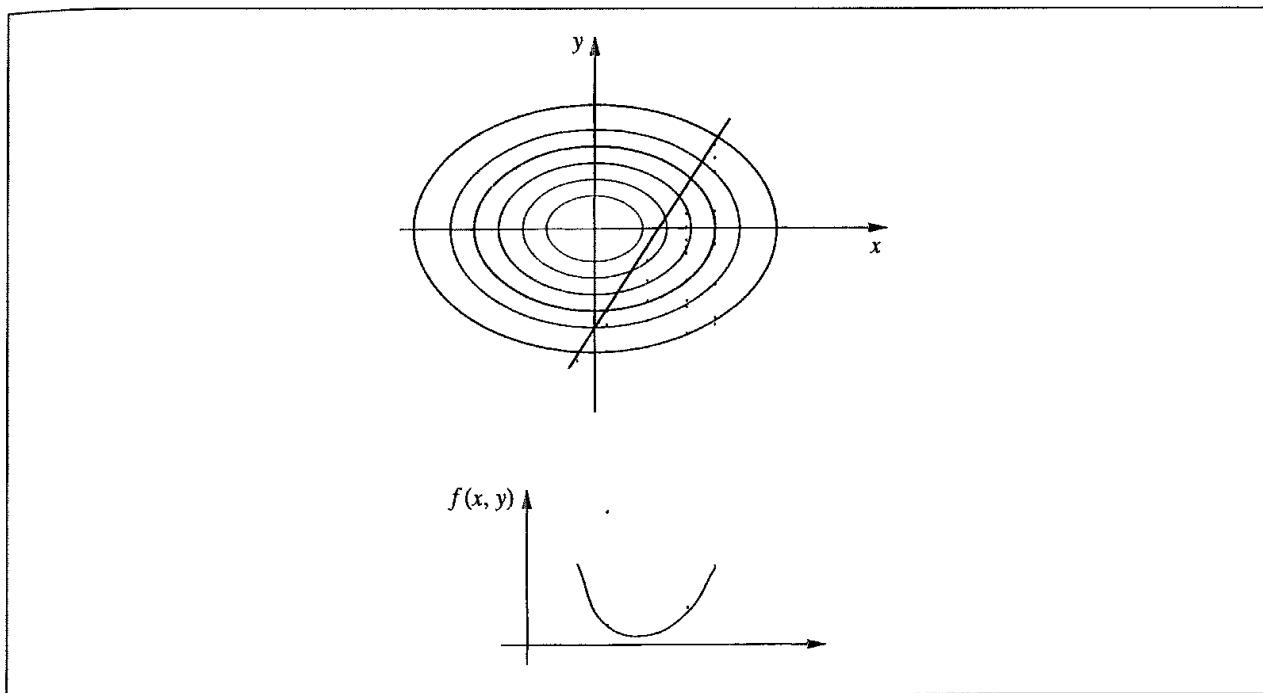


Fig 5.7 A line search is used to locate the minimum in the function in the direction of the gradient.

fitted function enables an approximation to the minimum along the line to be identified analytically. A new function can then be fitted to give a better estimate, as shown in Figure 5.8. Higher-order polynomials may give a better fit to the bracketing points but these can give incorrect interpolations when used with functions that change sharply in the bracketed region.

The gradient at the minimum point obtained from the line search will be perpendicular to the previous direction. Thus, when the line search method is used to locate the minimum along the gradient then the next direction in the steepest descents algorithm will be orthogonal to the previous direction (i.e. $\mathbf{g}_k \cdot \mathbf{g}_{k-1} = 0$).

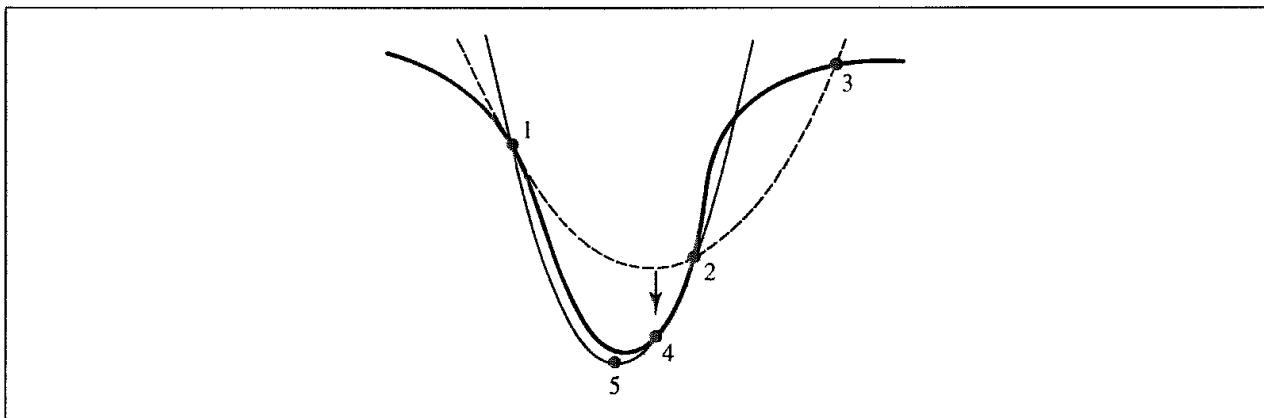


Fig 5.8: The minimum in a line search may be found more effectively by fitting an analytical function such as a quadratic to the initial set of three points (1, 2 and 3). A better estimate of the minimum can then be found by fitting a new function to the points 1, 2 and 4 and finding its minimum. (Figure adapted from Press W H, B P Flannery, S A Teukolsky and W T Vetterling 1992 Numerical Recipes in Fortran Cambridge, Cambridge University Press)

5.4.3 Arbitrary Step Approach

As the line search may itself be computationally demanding we could obtain the new coordinates by taking a step of arbitrary length along the gradient unit vector \mathbf{s}_k . The new set of coordinates after step k would then be given by the equation:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{s}_k \quad (5.5)$$

λ_k is the *step size*. In most applications of the steepest descents algorithm in molecular modelling the step size initially has a predetermined default value. If the first iteration leads to a reduction in energy, the step size is increased by a multiplicative factor (e.g. 1.2) for the second iteration. This process is repeated so long as each iteration reduces the energy. When a step produces an increase in energy, it is assumed that the algorithm has leapt across the valley which contains the minimum and up the slope on the opposite face. The step size is then reduced by a multiplicative factor (e.g. 0.5). The step size depends upon the nature of the energy surface; for a flat surface large step sizes would be appropriate but for a narrow, twisting gully a much smaller step would be more suitable. The arbitrary step method may require more steps to reach the minimum but it can often require fewer function evaluations (and thus less computer time) than the more rigorous line search approach.

The steepest descents method works as follows for our trial function, $f(x, y) = x^2 + 2y^2$. Differentiating the function gives $df = 2x \, dx + 4y \, dy$ and so the gradient at any point (x, y) equals $4y/2x$. The direction of the first move from the point $(9, 0, 9, 0)$ is $(-18, 0, -36, 0)$ and the equation of the line along which the search is performed is $y = 2x - 9$. The minimum of the function along this line can be obtained using Lagrange multipliers (see Section 1.10.5) and is at $(4.0, -1.0)$. The direction of the next move is the vector $(-8, 4)$ and the next line search is performed along the line $y = -0.5x + 1$. The minimum point along this line is $(2/3, 2/3)$ where the function has the value $4/3$. The third point found by the steepest descents method is at $(0.296, -0.074)$ where the function has the value 0.099. These moves are illustrated in Figure 5.9.

The direction of the gradient is determined by the largest interatomic forces and so steepest descents is a good method for relieving the highest-energy features in an initial configuration. The method is generally robust even when the starting point is far from a minimum, where the harmonic approximation to the energy surface is often a poor assumption. However, it suffers from the problem that many small steps will be performed when proceeding down a long narrow valley. The steepest descents method is forced to make a right-angled turn at each point, even though that might not be the best route to the minimum. The path oscillates and continually overcorrects itself, as illustrated in Figure 5.10; later steps reintroduce errors that were corrected by earlier moves.

5.4.4 Conjugate Gradients Minimisation

The conjugate gradients method produces a set of directions which does not show the oscillatory behaviour of the steepest descents method in narrow valleys. In the steepest descents method both the gradients and the direction of successive steps are orthogonal.

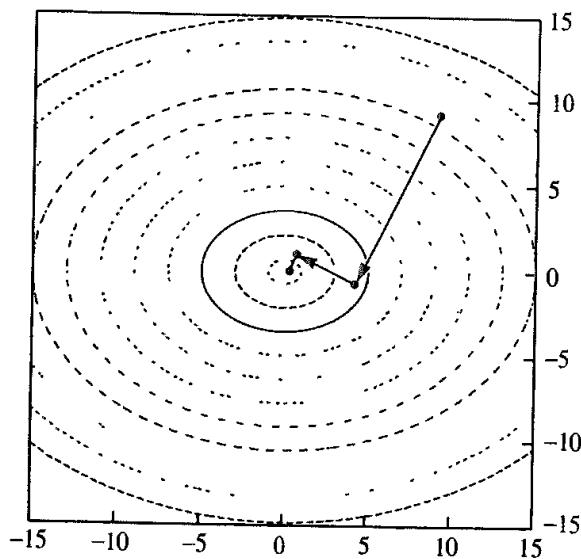


Fig 5.9. Application of steepest descents to the function $x^2 + 2y^2$.

In conjugate gradients, the gradients at each point are orthogonal but the directions are *conjugate* (indeed, the method is more properly called the conjugate directions method). A set of conjugate directions has the property that for a quadratic function of M variables, the minimum will be reached in M steps. The conjugate gradients method moves in a direction \mathbf{v}_k from point \mathbf{x}_k where \mathbf{v}_k is computed from the gradient at the point and the previous direction vector \mathbf{v}_{k-1} :

$$\mathbf{v}_k = -\mathbf{g}_k + \gamma_k \mathbf{v}_{k-1} \quad (5.6)$$

γ_k is a scalar constant given by

$$\gamma_k = \frac{\mathbf{g}_k \cdot \mathbf{g}_k}{\mathbf{g}_{k-1} \cdot \mathbf{g}_{k-1}} \quad (5.7)$$

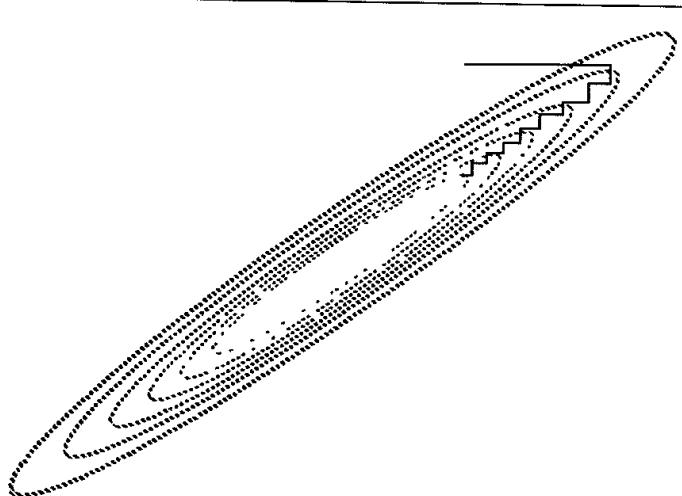


Fig 5.10 The steepest descents method can give undesirable behaviour in a long narrow valley

In the conjugate gradients method all of the directions and gradients satisfy the following relationships:

$$\mathbf{g}_i \cdot \mathbf{g}_j = 0 \quad (5.8)$$

$$\mathbf{v}_i \cdot \nabla''_{ij} \cdot \mathbf{v}_j = 0 \quad (5.9)$$

$$\mathbf{g}_i \cdot \mathbf{v}_j = 0 \quad (5.10)$$

Clearly Equation (5.6) can only be used from the second step onwards and so the first step in the conjugate gradients method is the same as the steepest descents (i.e. in the direction of the gradient). The line search method should ideally be used to locate the one-dimensional minimum in each direction to ensure that each gradient is orthogonal to all previous gradients and that each direction is conjugate to all previous directions. However, an arbitrary step method is also possible.

The conjugate gradients method deals with our simple quadratic function $f(x, y) = x^2 + 2y^2$ as follows. From the initial point $(9, 9)$ we move to the same point as in steepest descents, $(4, -1)$. To find the direction of the next move, we first determine the negative gradient at the current point. This is the vector $(-8, 4)$. This is then combined with the vector corresponding to minus the gradient at the initial point, $(-18, -36)$ multiplied by γ :

$$\mathbf{v}_k = \begin{pmatrix} -8 \\ 4 \end{pmatrix} + \frac{(-8)^2 + (4)^2}{(-18)^2 + (-36)^2} \begin{pmatrix} -18 \\ -36 \end{pmatrix} = \begin{pmatrix} -80/9 \\ +20/9 \end{pmatrix} \quad (5.11)$$

To locate the second point we therefore need to perform a line search along the line with gradient $-1/4$ that passes through the point $(4, -1)$. The minimum along this line is at the origin, at the true minimum of the function. The conjugate gradients method thus locates the exact minimum of the function exactly in just two moves, as illustrated in Figure 5.11.

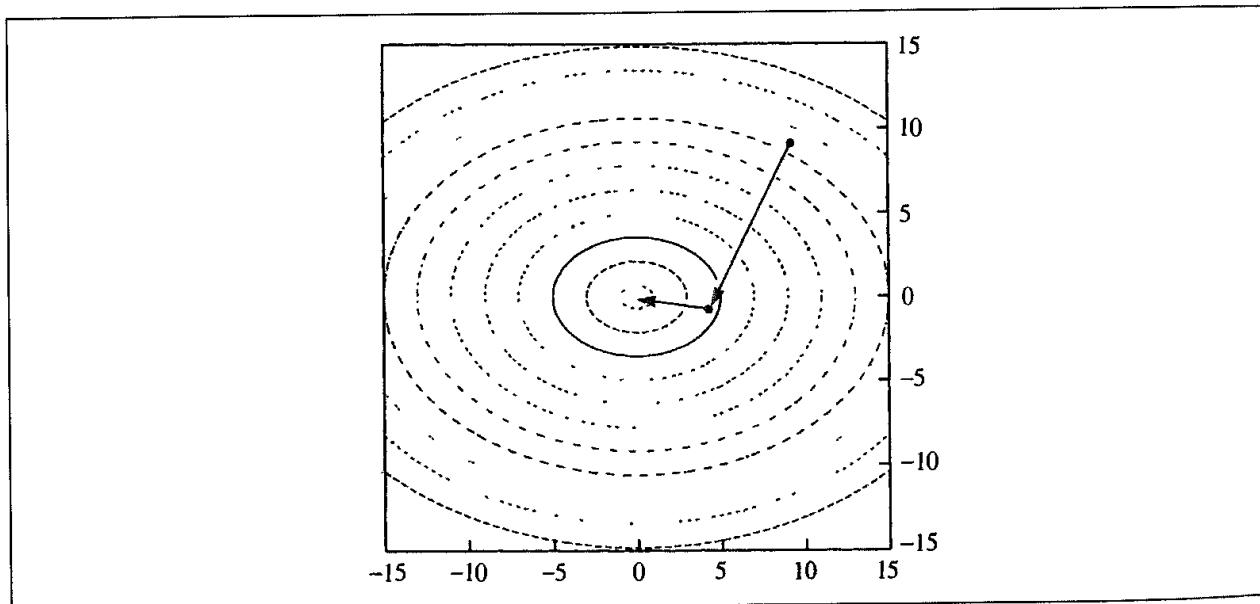


Fig 5.11 Application of conjugate gradients method to the function $x^2 + 2y^2$.

Several variants of the conjugate gradients method have been proposed. The formulation given in Equation (5.7) is the original Fletcher–Reeves algorithm. Polak and Ribiere proposed an alternative form for the scalar constant γ_k :

$$\gamma_k = \frac{(\mathbf{g}_k - \mathbf{g}_{k-1}) \cdot \mathbf{g}_k}{\mathbf{g}_{k-1} \cdot \mathbf{g}_{k-1}} \quad (5.12)$$

For a purely quadratic function the Polak–Ribiere method is identical to the Fletcher–Reeves algorithm as all gradients will be orthogonal. However, most functions of interest, including those used in molecular modelling, are at best only approximately quadratic. Polak and Ribiere claimed that their method performed better than the original Fletcher–Reeves algorithm, at least for the functions that they examined.

5.5 Second Derivative Methods: The Newton–Raphson Method

Second-order methods use not only the first derivatives (i.e. the gradients) but also the second derivatives to locate a minimum. Second derivatives provide information about the curvature of the function. The *Newton–Raphson* method is the simplest second-order method. Recall our Taylor series expansion about the point x_k , Equation (5.2):

$$\mathcal{V}(x) = \mathcal{V}(x_k) + (x - x_k)\mathcal{V}'(x_k) + (x - x_k)^2\mathcal{V}''(x_k)/2 + \dots \quad (5.13)$$

The first derivative of $\mathcal{V}(x)$ is:

$$\mathcal{V}'(x) = x\mathcal{V}'(x_k) + (x - x_k)\mathcal{V}''(x_k) \quad (5.14)$$

If the function is purely quadratic, the second derivative is the same everywhere, and so $\mathcal{V}''(x) = \mathcal{V}''(x_k)$.

At the minimum ($x = x^*$) $\mathcal{V}'(x^*) = 0$ and so

$$x^* = x_k - \mathcal{V}'(x_k)/\mathcal{V}''(x_k) \quad (5.15)$$

For a multidimensional function: $\mathbf{x}^* = \mathbf{x}_k - \mathcal{V}'(\mathbf{x}_k)\mathcal{V}''^{-1}(\mathbf{x}_k)$.

$\mathcal{V}''^{-1}(\mathbf{x}_k)$ is the inverse Hessian matrix of second derivatives, which, in the Newton–Raphson method, must therefore be inverted. This can be computationally demanding for systems with many atoms and can also require a significant amount of storage. The Newton–Raphson method is thus more suited to small molecules (usually less than 100 atoms or so). For a purely quadratic function the Newton–Raphson method finds the minimum in one step from any point on the surface, as we will now show for our function $f(x, y) = x^2 + 2y^2$.

The Hessian matrix for this function is:

$$\mathbf{f}'' = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \quad (5.16)$$

The inverse of this matrix is:

$$\mathbf{f}''^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \quad (5.17)$$

The minimum is obtained using Equation (5.15):

$$\mathbf{x}^* = \begin{pmatrix} 9 \\ 9 \end{pmatrix} - \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 18 \\ 36 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5.18)$$

In practice, of course, the surface is only quadratic to a first approximation and so a number of steps will be required, at each of which the Hessian matrix must be calculated and inverted. The Hessian matrix of second derivatives must be *positive definite* in a Newton-Raphson minimisation. A positive definite matrix is one for which all the eigenvalues are positive. When the Hessian matrix is not positive definite then the Newton-Raphson method moves to points (e.g. saddle points) where the energy increases. In addition, far from a minimum the harmonic approximation is not appropriate and the minimisation can become unstable. One solution to this problem is to use a more robust method to get near to the minimum (i.e. where the Hessian is positive definite) before applying the Newton-Raphson method.

5.5.1 Variants on the Newton–Raphson Method

There are a number of variations on the Newton–Raphson method, many of which aim to eliminate the need to calculate the full matrix of second derivatives. In addition, a family of methods called the quasi-Newton methods require only first derivatives and gradually construct the inverse Hessian matrix as the calculation proceeds. One simple way in which it may be possible to speed up the Newton–Raphson method is to use the same Hessian matrix for several successive steps of the Newton–Raphson algorithm with only the gradients being recalculated at each iteration.

A widely used algorithm is the *block-diagonal Newton–Raphson* method in which just one atom is moved at each iteration. Consequently all terms of the form $\partial^2\mathcal{V}/\partial x_i \partial x_j$, where i and j refer to the Cartesian coordinates of atoms other than the atom being moved, will be zero. This only leaves those terms which involve the coordinates of the atom being moved and so reduces the problem to the trivial one of inverting a 3×3 matrix. However, the block-diagonal approach can be less efficient when the motions of some atoms are closely coupled, such as the concerted movements of connected atoms in a phenyl ring.

5.6 Quasi-Newton Methods

Calculation of the inverse Hessian matrix can be a potentially time-consuming operation that represents a significant drawback to the ‘pure’ second derivative methods such as Newton–Raphson. Moreover, one may not be able to calculate analytical second derivatives, which are preferable. The quasi-Newton methods (also known as variable metric methods) gradually build up the inverse Hessian matrix in successive iterations. That is, a sequence of

matrices \mathbf{H}_k is constructed that has the property

$$\lim_{k \rightarrow \infty} \mathbf{H}_k = \mathcal{V}^{\prime\prime-1} \quad (5.19)$$

At each iteration k , the new positions \mathbf{x}_{k+1} are obtained from the current positions \mathbf{x}_k , the gradient \mathbf{g}_k and the current approximation to the inverse Hessian matrix \mathbf{H}_k :

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k \mathbf{g}_k \quad (5.20)$$

This formula is exact for a quadratic function, but for 'real' problems a line search may be desirable. This line search is performed along the vector $\mathbf{x}_{k+1} - \mathbf{x}_k$. It may not be necessary to locate the minimum in the direction of the line search very accurately, at the expense of a few more steps of the quasi-Newton algorithm. For quantum mechanics calculations the additional energy evaluations required by the line search may prove more expensive than using the more approximate approach. An effective compromise is to fit a function to the energy and gradient at the current point \mathbf{x}_k and at the point \mathbf{x}_{k+1} and determine the minimum in the fitted function.

Having moved to the new positions \mathbf{x}_{k+1} , \mathbf{H} is updated from its value at the previous step according to a formula depending upon the specific method being used. The methods of Davidon-Fletcher-Powell (DFP), Broyden-Fletcher-Goldfarb-Shanno (BFGS) and Murtaugh-Sargent (MS) are commonly encountered, but there are many others. These methods converge to the minimum, for a quadratic function of M variables, in M steps. The DFP formula is:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{(\mathbf{x}_{k+1} - \mathbf{x}_k) \otimes (\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k) \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} - \frac{[\mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)] \otimes [\mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)]}{(\mathbf{g}_{k+1} - \mathbf{g}_k) \cdot \mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} \quad (5.21)$$

The symbol \otimes when interposed between two vectors means that a matrix is to be formed. The ij th element of the matrix $\mathbf{u} \otimes \mathbf{v}$ is obtained by multiplying \mathbf{u}_i by \mathbf{v}_j .

The BFGS formula differs from the DFP equation by an additional term:

$$\begin{aligned} \mathbf{H}_{k+1} = \mathbf{H}_k &+ \frac{(\mathbf{x}_{k+1} - \mathbf{x}_k) \otimes (\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k) \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} - \frac{[\mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)] \otimes [\mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)]}{(\mathbf{g}_{k+1} - \mathbf{g}_k) \cdot \mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} \\ &+ [(\mathbf{g}_{k+1} - \mathbf{g}_k) \cdot \mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)] \mathbf{u} \otimes \mathbf{u} \end{aligned} \quad (5.22)$$

where

$$\mathbf{u} = \frac{(\mathbf{x}_{k+1} - \mathbf{x}_k)}{(\mathbf{x}_{k+1} - \mathbf{x}_k) \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} - \frac{[\mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)]}{(\mathbf{g}_{k+1} - \mathbf{g}_k) \cdot \mathbf{H}_k \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} \quad (5.23)$$

The MS formula is:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{[(\mathbf{x}_{k+1} - \mathbf{x}_k) - \mathbf{H}_k(\mathbf{g}_{k+1} - \mathbf{g}_k)] \otimes [(\mathbf{x}_{k+1} - \mathbf{x}_k) - \mathbf{H}_k(\mathbf{g}_{k+1} - \mathbf{g}_k)]}{[(\mathbf{x}_{k+1} - \mathbf{x}_k) - \mathbf{H}_k(\mathbf{g}_{k+1} - \mathbf{g}_k)] \cdot (\mathbf{g}_{k+1} - \mathbf{g}_k)} \quad (5.24)$$

All of these methods use just the new and current points to update the inverse Hessian. The default algorithm used in the Gaussian series of molecular orbital programs [Schlegel 1982] makes use of more of the previous points to construct the Hessian (and thence the inverse Hessian), giving better convergence properties. Another feature of this method is its use

of a quartic polynomial that is guaranteed to have just one local minimum in the line search. The DFP, BFGS and MS methods can also be used with numerical derivatives, but alternative approaches may be more effective under such circumstances.

The matrix \mathbf{H} is often initialised to the unit matrix \mathbf{I} . The performance of the quasi-Newton algorithms can be improved by using a better estimate of the inverse Hessian than just the unit matrix. The unit matrix gives no information about the bonding in the system, nor does it identify any coupling between the various degrees of freedom. For example, a molecular mechanics calculation can be used to provide an initial guess to \mathbf{H} prior to a quantum mechanical calculation. Alternatively the matrix can be obtained from a quantum mechanical calculation at a lower level of theory (e.g. semi-empirical or with a smaller basis set).

5.7 Which Minimisation Method Should I Use?

The choice of minimisation algorithm is dictated by a number of factors, including the storage and computational requirements, the relative speeds with which the various parts of the calculation can be performed, the availability of analytical derivatives and the robustness of the method. Thus, any method that requires the Hessian matrix to be stored (let alone its inverse calculated) may present memory problems when applied to systems containing thousands of atoms. Calculations on systems of this size are invariably performed using molecular mechanics, and so the steepest descents and the conjugate gradients methods are very popular here. For molecular mechanics calculations on small molecules, the Newton-Raphson method may be used, although this algorithm can have problems with structures that are far from a minimum. For this reason it is usual to perform a few steps of minimisation using a more robust method such as the simplex or steepest descents before applying the Newton-Raphson algorithm. Analytical expressions for both first and second derivatives are available for most of the terms found in common force fields.

The performance of the steepest descents and conjugate gradients methods is contrasted in the following example. A model of the antibiotic netropsin (Figure 5.12) bound to DNA was constructed using an automated docking program. This initial model was then subjected to two stages of minimisation. In the first stage, the aim was to produce a structure that did not

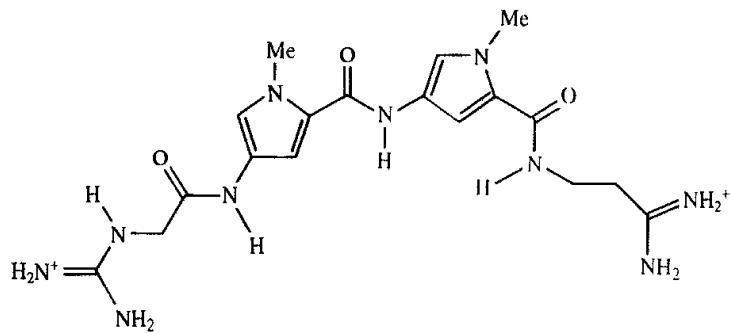


Fig 5.12: The DNA inhibitor netropsin

Method	Initial refinement (Av. gradient <1 kcal Å⁻²)		Stringent minimisation (Av. gradient <0.1 kcal Å⁻²)	
	CPU time (s)	Number of iterations	CPU time (s)	Number of iterations
Steepest descents	67	98	1405	1893
Conjugate gradients	149	213	257	367

Table 5.1 A comparison of the steepest descents and conjugate gradients methods for an initial refinement and a stringent minimisation.

have any significant high-energy interactions. The structure was then further minimised to give a structure much closer to the minimum. The results are shown in Table 5.1.

This study shows that the steepest descent method can actually be superior to conjugate gradients when the starting structure is some way from the minimum. However, conjugate gradients is much better once the initial strain has been removed.

Quantum mechanical calculations are restricted to systems with relatively small numbers of atoms, and so storing the Hessian matrix is not a problem. As the energy calculation is often the most time-consuming part of the calculation, it is desirable that the minimisation method chosen takes as few steps as possible to reach the minimum. For many levels of quantum mechanics theory analytical first derivatives are available. However, analytical second derivatives are only available for a few levels of theory and can be expensive to compute. The quasi-Newton methods are thus particularly popular for quantum mechanical calculations.

When using internal coordinates in a quantum mechanical minimisation it can be important to use an appropriate Z-matrix as input. For many systems the Z-matrix can often be written in many different ways as there are many combinations of internal coordinates. There should be no strong coupling between the coordinates. *Dummy atoms* can often help in the construction of an appropriate Z-matrix. A dummy atom is used solely to define the geometry and has no nuclear charge and no basis functions. A simple example of the use of dummy atoms is for a linear molecule such as HN_3 , where the angle of 180° would cause problems. The geometry of this molecule can be defined using a dummy atom as illustrated in Figure 5.13; the associated Z-matrix for this system would be:

1	N						
2	N	1	RN1N2				
3	X	1	1.0	2	90.0		
4	N	1	RN1N4	3	AN4N1X	2	180.0
5	H	4	RN4H	1	AHN4N1	3	180.0

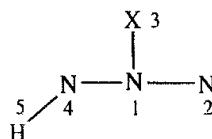


Fig 5.13 Internal coordinates of HN_3 molecule defined using dummy atom X

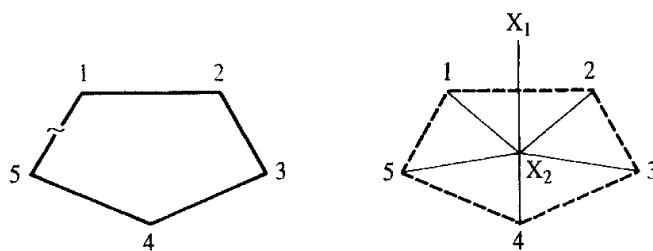


Fig. 5.14: The ring closure bond between atoms 1 and 5 would be strongly coupled to the other internal coordinates (left) unless dummy atoms are used to define the Z-matrix (right)

Strong coupling between coordinates can give long ‘valleys’ in the energy surface, which may also present problems. Care must be taken when defining the Z-matrix for cyclic systems in particular. The natural way to define a cyclic compound would be to number the atoms sequentially around the ring. However, this would then mean that the ring closure bond will be very strongly coupled to all of the other bonds, angles and torsion angles (Figure 5.14). A better definition uses a dummy atom placed at the centre of the ring (Figure 5.14). Some quantum mechanics programs are able to convert the input coordinates (be they Cartesian or internal) into the most efficient set for minimisation so removing from the user the problems of trying to decide what is an appropriate set of internal coordinates. For energy minimisations redundant internal coordinates have been shown to give significant improvements in efficiency compared with Cartesian coordinates or non-redundant internal coordinates, especially for flexible and polycyclic systems [Peng *et al.* 1996]. The redundant internal coordinates employed generally comprise the bond lengths, angles and torsion angles in the system. These methods obviously also require the means to interconvert between the internal coordinate representation and the Cartesian coordinates that are often used as input and desired as output. Of particular importance is the need to transform energy derivatives and the Hessian matrices (if appropriate).

5.7.1 Distinguishing Between Minima, Maxima and Saddle Points

A configuration at which all the first derivatives are zero need not necessarily be a minimum point; this condition holds at both maxima and saddle points as well. From simple calculus we know that the second derivative of a function of one variable, $f'(x)$ is positive at a minimum and negative at a maximum. It is necessary to calculate the eigenvalues of the Hessian matrix to distinguish between minima, maxima and saddle points. At a minimum point there will be six zero and $3N - 6$ positive eigenvalues if $3N$ Cartesian coordinates are used. The six zero eigenvalues correspond to the translational and rotational degrees of freedom of the molecule (thus these six zero eigenvalues are not obtained when internal coordinates are used). At a maximum point all eigenvalues are negative and at a saddle point one or more eigenvalues are negative. We will consider the uses of the eigenvalue and eigenvector information in Sections 5.8 and 5.9.

5.7.2 Convergence Criteria

In contrast to the simple analytical functions that we have used to illustrate the operation of the various minimisation methods, in ‘real’ molecular modelling applications it is rarely possible to identify the ‘exact’ location of minima and saddle points. We can only ever hope to find an approximation to the true minimum or saddle point. Unless instructed otherwise, most minimisation methods would keep going forever, moving ever closer to the minimum. It is therefore necessary to have some means to decide when the minimisation calculation is sufficiently close to the minimum and so can be terminated. Any calculation is of course limited by the precision with which numbers can be stored on the computer, but in most instances it is usual to stop well before this limit is reached. A simple strategy is to monitor the energy from one iteration to the next and to stop when the difference in energy between successive steps falls below a specified threshold. An alternative is to monitor the change in coordinates and to stop when the difference between successive configurations is sufficiently small. A third method is to calculate the root-mean-square gradient. This is obtained by adding the squares of the gradients of the energy with respect to the coordinates, dividing by the number of coordinates and taking the square root:

$$\text{RMS} = \sqrt{\frac{\mathbf{g}^T \mathbf{g}}{3N}} \quad (5.25)$$

It is also useful to monitor the maximum value of the gradient to ensure that the minimisation has properly relaxed all the degrees of freedom and has not left a large amount of strain in one or two coordinates.

5.8 Applications of Energy Minimisation

Energy minimisation is very widely used in molecular modelling and is an integral part of techniques such as conformational search procedures (Chapter 9). Energy minimisation is also used to prepare a system for other types of calculation. For example, energy minimisation may be used prior to a molecular dynamics or Monte Carlo simulation in order to relieve any unfavourable interactions in the initial configuration of the system. This is especially recommended for simulations of complex systems such as macromolecules or large molecular assemblies. In the following sections we will discuss some techniques that are specifically associated with energy minimisation methods.

5.8.1 Normal Mode Analysis

The molecular mechanics or quantum mechanics energy at an energy minimum corresponds to a hypothetical, motionless state at 0 K. Experimental measurements are made on molecules at a finite temperature when the molecules undergo translational, rotational and vibration motion. To compare the theoretical and experimental results it is

necessary to make appropriate corrections to allow for these motions. These corrections are calculated using standard statistical mechanics formulae. The internal energy $U(T)$ at a temperature T is given by:

$$U(T) = U_{\text{trans}}(T) + U_{\text{rot}}(T) + U_{\text{vib}}(T) + U_{\text{vib}}(0) \quad (5.26)$$

If all translational and rotational modes are fully accessible in accordance with the equipartition theorem, then $U_{\text{trans}}(T)$ and $U_{\text{rot}}(T)$ are both equal to $\frac{3}{2}k_B T$ per molecule (except that $U_{\text{rot}}(T)$ equals $k_B T$ for a linear molecule); k_B is Boltzmann's constant. However, the vibrational energy levels are often only partially excited at room temperature. The vibrational contribution to the internal energy at a temperature T thus requires knowledge of the actual vibrational frequencies. The vibrational contribution equals the difference in the vibrational enthalpy at the temperature T and at 0 K and is given by:

$$U_{\text{vib}}(T) = \sum_{i=1}^{N_{\text{nm}}} \left(\frac{h\nu_i}{2} + \frac{h\nu_i}{\exp[h\nu_i/k_B T] + 1} \right) \quad (5.27)$$

N_{nm} is the number of *normal vibrational modes* for the system. Even the zero-point energy ($U_{\text{vib}}(0)$, obtained by summing $\frac{1}{2}h\nu_i$ for each normal mode) can be quite substantial, amounting to about 100 kcal/mol for a six-carbon alkane. Other thermodynamic quantities such as entropies and free energies may also be calculated from the vibrational frequencies using the relevant statistical mechanics expressions.

Normal modes are useful because they correspond to collective motions of the atoms in a coupled system that can be individually excited. The three normal modes of water are schematically illustrated in Figure 5.15; a non-linear molecule with N atoms has $3N - 6$ normal modes. The frequencies of the normal modes together with the displacements of the individual atoms may be calculated from a molecular mechanics force field or from the wavefunction using the Hessian matrix of second derivatives (\mathcal{V}''). Of course, if we have used an appropriate minimisation algorithm then we already know the Hessian. The Hessian must first be converted to the equivalent force-constant matrix in *mass-weighted coordinates* (F), as follows.

$$F = M^{-1/2} \mathcal{V}'' M^{-1/2} \quad (5.28)$$

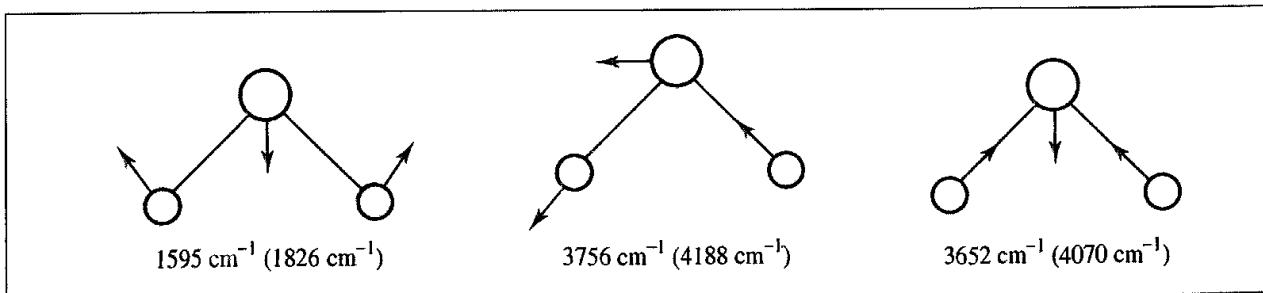


Fig. 5.15 Normal modes of water. Experimental and (calculated) frequencies are shown. Theoretical frequencies calculated using a 6-31G* basis set

M is a diagonal matrix of dimension $3N \times 3N$, containing the atomic masses. All elements of M are zero except those on the diagonal; $M_{1,1} = m_1$, $M_{2,2} = m_1$, $M_{3,3} = m_1$, $M_{4,4} = m_2, \dots, M_{3N-2,3N-2} = m_N$, $M_{3N-1,3N-1} = m_N$, $M_{3N,3N} = m_N$. Each non-zero element of $M^{-1/2}$ is thus the inverse square root of the mass of the appropriate atom. The masses of the atoms must be taken into account because a force of a given magnitude will have a different effect upon a larger mass than a smaller one. For example, the force constant for a bond to a deuterium atom is, to a good approximation, the same as to a proton, yet the different mass of the deuteron gives a different motion and a different zero-point energy. The use of mass-weighted coordinates takes care of these problems.

We next solve the secular equation $|\mathbf{F} - \mathbf{I}| = 0$ to obtain the eigenvalues and eigenvectors of the matrix \mathbf{F} . This step is usually performed using matrix diagonalisation, as outlined in Section 1.10.3. If the Hessian is defined in terms of Cartesian coordinates then six of these eigenvalues will be zero as they correspond to translational and rotational motion of the entire system. The frequency of each normal mode is then calculated from the eigenvalues using the relationship:

$$\nu_i = \frac{\sqrt{\lambda_i}}{2\pi} \quad (5.29)$$

As a simple example of a normal mode calculation consider the linear triatomic system in Figure 5.16. We shall just consider motion along the long axis of the molecule. The displacements of the atoms from their equilibrium positions along this axis are denoted by ξ_i . It is assumed that the displacements are small compared with the equilibrium values l_0 and the system obeys Hooke's law with bond force constants k . The potential energy is given by:

$$\mathcal{V} = \frac{1}{2}k(\xi_1 - \xi_2)^2 + \frac{1}{2}k(\xi_2 - \xi_3)^2 \quad (5.30)$$

We next calculate the first and then the second derivatives of the potential energy with respect to the three coordinates ξ_1 , ξ_2 and ξ_3 :

$$\frac{\partial \mathcal{V}}{\partial \xi_1} = k(\xi_1 - \xi_2); \quad \frac{\partial \mathcal{V}}{\partial \xi_2} = -k(\xi_1 - \xi_2) + k(\xi_2 - \xi_3); \quad \frac{\partial \mathcal{V}}{\partial \xi_3} = -k(\xi_2 - \xi_3) \quad (5.31)$$

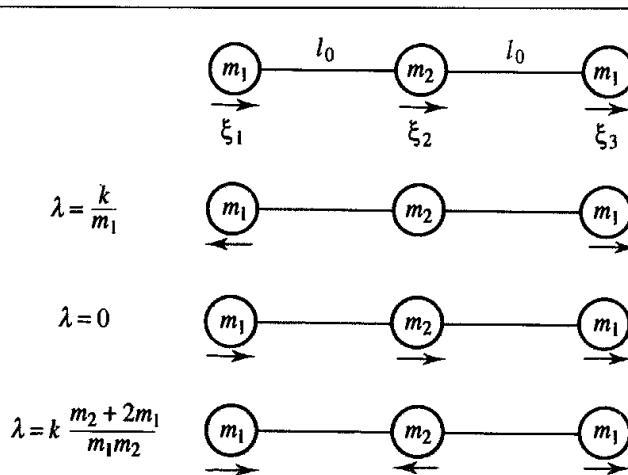


Fig 5.16. Linear three-atom system with results of normal mode calculation

The second derivatives are conveniently represented as a 3×3 matrix:

$$\begin{vmatrix} k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{vmatrix} \quad (5.32)$$

The mass-weighted matrix is

$$\begin{vmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{vmatrix} \quad (5.33)$$

The secular equation to be solved is thus:

$$\begin{vmatrix} \frac{k}{m_1} - \lambda & -\frac{k}{\sqrt{m_1}\sqrt{m_2}} & 0 \\ -\frac{k}{\sqrt{m_1}\sqrt{m_2}} & \frac{2k}{m_2} - \lambda & -\frac{k}{\sqrt{m_1}\sqrt{m_2}} \\ 0 & -\frac{k}{\sqrt{m_1}\sqrt{m_2}} & \frac{k}{m_1} - \lambda \end{vmatrix} = 0 \quad (5.34)$$

This determinant leads to a cubic in λ which has three roots (λ_k), each corresponding to a different mode of motion:

$$\lambda = \frac{k}{m_1}, \quad \lambda = 0, \quad \lambda = k \frac{m_2 + 2m_1}{m_1 m_2} \quad (5.35)$$

The corresponding frequencies can be obtained from Equation (5.29). The amplitudes (A) of each normal mode are given by the eigenvector solutions of the secular equation $\mathbf{F}\mathbf{A} = \lambda\mathbf{A}$. If A_1 , A_2 and A_3 are the amplitudes of each atom then the amplitudes obtained for each eigenvalue are:

$$\lambda = \frac{k}{m_1} : \quad A_1 = -A_3; \quad A_2 = 0 \quad (5.36)$$

$$\lambda = 0: \quad A_1 = A_3; \quad A_2 = \sqrt{\frac{m_2}{m_1}} A_1 \quad (5.37)$$

$$\lambda = k \frac{m_2 + 2m_1}{m_1 m_2} : \quad A_1 = A_3; \quad A_2 = -2\sqrt{\frac{m_1}{m_2}} A_1 \quad (5.38)$$

These normal modes are schematically illustrated in Figure 5.16. They correspond to a symmetric stretch, a translation and an asymmetric stretch respectively.

We have already seen how the results of normal mode calculations can be used to calculate thermodynamic quantities. The frequencies themselves can also be compared with the results of spectroscopic experiments, information which can be used in the parametrisation of a force field. For example, the experimental frequencies for the normal modes of water are shown in Figure 5.15, together with the frequencies determined using a 6-31G* *ab initio* calculation. The calculated values clearly deviate from those obtained experimentally, but the ratio of the experimental and theoretical frequencies is

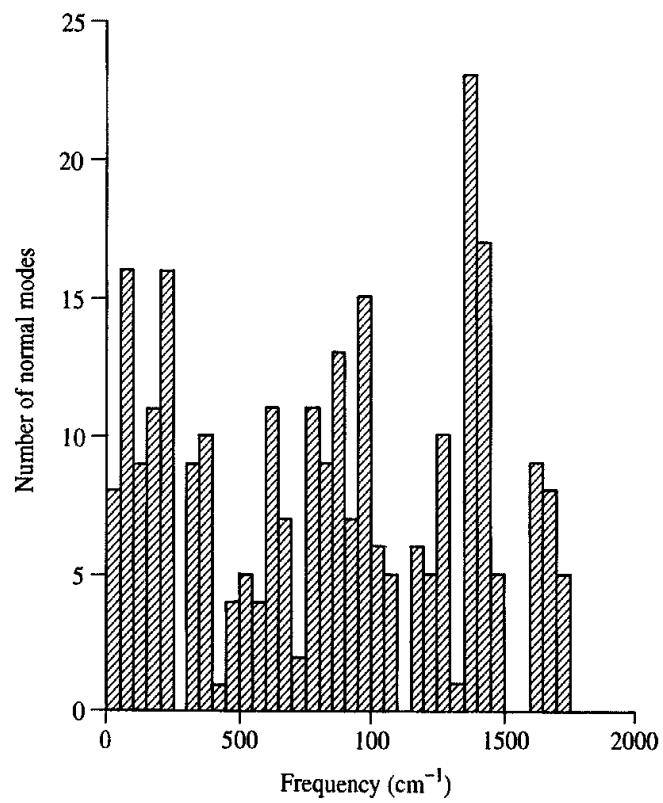
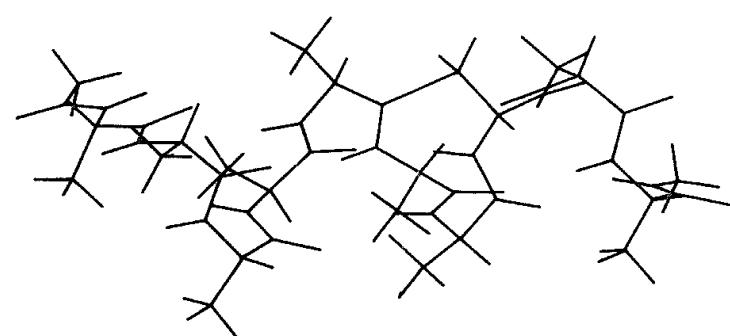


Fig. 5.17. Histogram of the normal modes calculated for a polyalanine polypeptide in an α -helical conformation. The height of each bar indicates the number of normal modes in each 50 cm^{-1} section.

remarkably consistent (at about 1.1). Such empirical scaling factors have been derived which enable frequencies obtained using a given level of theory to be converted to values for experiment or a higher level of theory [Pople *et al.* 1993]. The normal modes of much larger molecules can be calculated using molecular mechanics. For example, the vibrations of a helical polypeptide constructed from a sequence of ten alanine residues (112 atoms) are shown in Figure 5.17. In such cases it is usually the low-frequency vibrations that are of most interest as these correspond to the large-scale conformational motions of the molecule. The results of such analyses can be compared with molecular dynamics simulations from which vibrational contributions can also be extracted [Brooks and Karplus 1983].

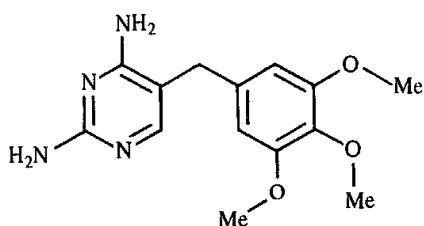


Fig. 5.18: Trimethoprim.

A normal mode calculation is based upon the assumption that the energy surface is quadratic in the vicinity of the energy minimum (the harmonic approximation). Deviations from the harmonic model can require corrections to calculated thermodynamic properties. One way to estimate anharmonic corrections is to calculate a force constant matrix using the atomic motions obtained from a molecular dynamics simulation; such simulations are not restricted to movements on a harmonic energy surface. The eigenvalues and eigenvectors are then calculated for this quasi-harmonic force-constant matrix in the normal way, giving a model which implicitly incorporates the anharmonic effects.

The harmonic approximation to the energy surface is found to be appropriate for well-defined energy minima such as the intramolecular degrees of freedom of small molecules and for some small intermolecular complexes. For larger systems such as liquids and large, 'floppy' molecules, the harmonic approximation breaks down. Such systems also have an extraordinarily large number of 'minima' on the energy surface. In such cases it is not possible to calculate accurately thermodynamic properties using energy minimisation and normal mode calculations. Rather, molecular dynamics or Monte Carlo simulations must be used to sample the energy surface from which properties can be derived, as we will discuss in Chapters 6–8.

5.8.2 The Study of Intermolecular Processes

One example of the use of minimisation methods and normal-mode analysis is the study by Hagler and co-workers of the binding of the antibacterial drug trimethoprim (Figure 5.18) to the enzyme dihydrofolate reductase (DHFR) [Dauber-Osguthorpe *et al.* 1988; Fisher *et al.* 1991]. DHFR catalyses the reduction of folic acid and dihydrofolic acid to tetrahydrofolic acid (Figure 5.19) and plays a vital metabolic role in the biosynthesis of nucleic acids in bacteria, protozoa, plants and animals. Trimethoprim exploits the structural differences between bacterial and vertebrate DHFR, binding much more strongly to the former, and is clinically used as an antibacterial agent. Inhibitors of human DHFR are used in cancer therapy. Hagler and colleagues applied energy minimisation to an isolated trimethoprim molecule, to the crystal structure of trimethoprim, to trimethoprim in the presence of water molecules, and to trimethoprim in intermolecular complexes with DHFR from both bacterial and vertebrate sources. An important observation was that the conformation of the trimethoprim, when bound to the enzyme, was significantly different from that obtained for the isolated molecule. This reinforces the view that the use of

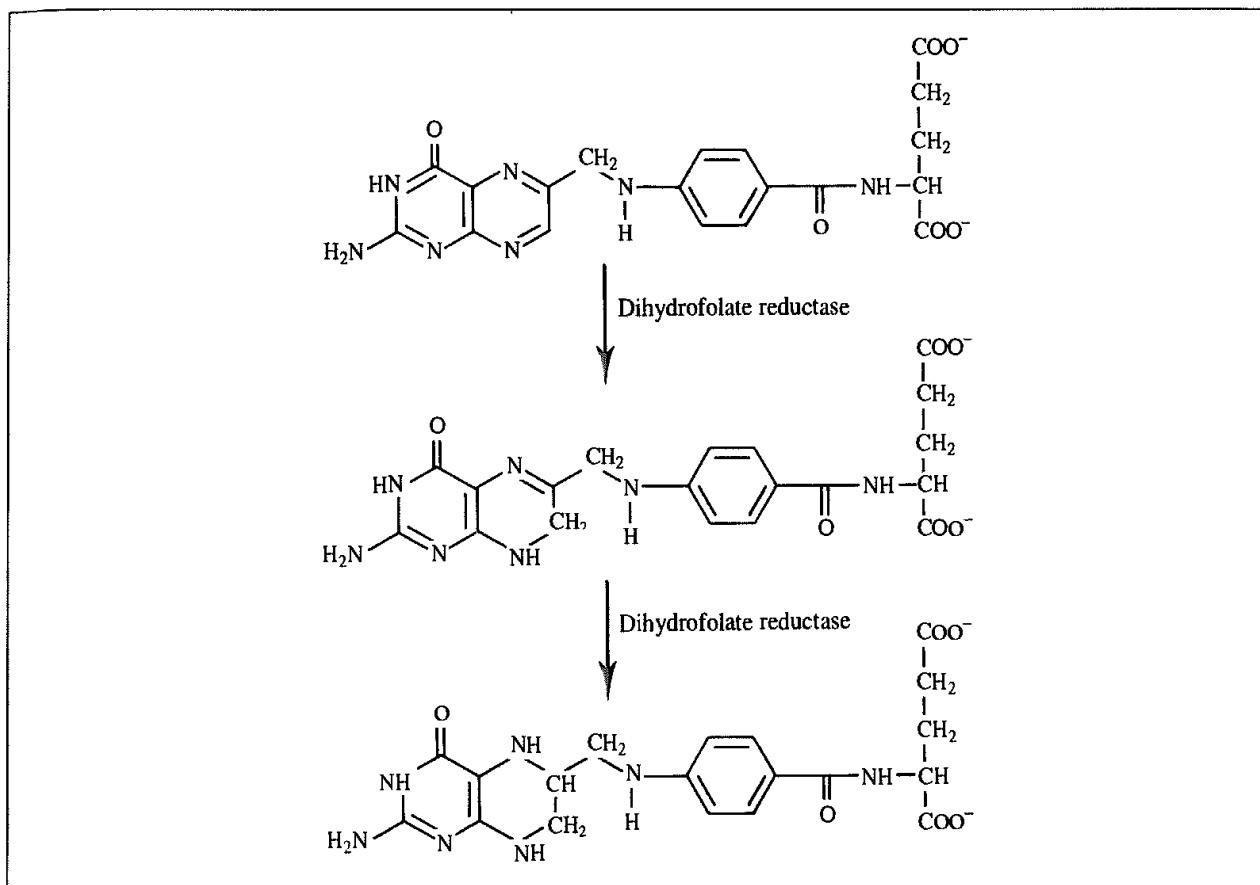


Fig 5.19: DHFR catalyses the reduction of folic acid to tetrahydrofolic acid.

structures obtained from energy minimisation calculations on isolated molecules can lead to misleading conclusions. Intermolecular interactions with the receptor enable the ligand to adopt a conformation whose intramolecular energy is significantly higher than any of its minimum energy structures.

A normal mode analysis on the isolated and bound trimethoprim molecules enabled an estimate to be made of the entropic contribution to binding. Low-frequency modes for the isolated ligand were found to be shifted to higher frequencies for the ligand in the enzyme complex, reflecting a restriction of the motion of the ligand by the protein. This entropic contribution to the free energy of binding was predicted to be quite significant, indicating that conclusions based solely upon energies may be misleading.

5.9 Determination of Transition Structures and Reaction Pathways

Chemists are interested not only in the thermodynamics of a process (the relative stability of the various species) but also in its kinetics (the rate of conversion from one structure to another). Knowledge of the minimum points on an energy surface enables thermodynamic data to be interpreted, but for the kinetics it is necessary to investigate the nature of the

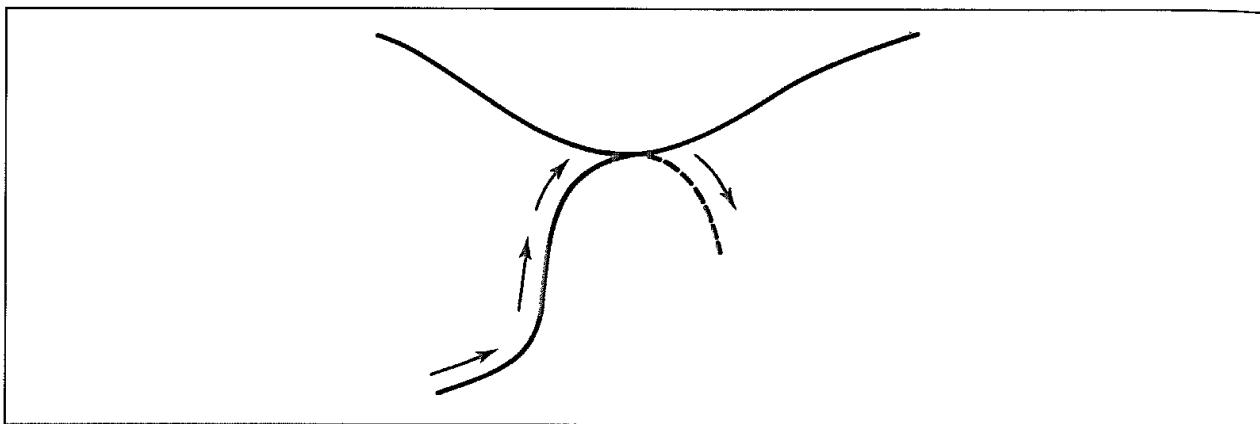


Fig. 5.20: The lowest-energy path from one minimum to another passes through a saddle point

energy surface away from the minimum points. In particular, we would like to know how the system changes from one minimum to another. What changes in geometry are involved, and how does the energy vary during the transition? The minimum points on the energy surface may be the reactants and products of a chemical reaction, two conformations of a molecule, or two molecules that associate to form a non-covalently bound bimolecular complex. We shall use the term 'reaction pathway' to describe the path between two minima, but our use of the word 'reaction' does not necessarily mean that bond making and/or breaking is involved. Many methods have been proposed for finding transition structures and elucidating reaction pathways. We do not have space to cover all of the methods, and so we shall restrict our discussion to some of the more common approaches.

As a system moves from one minimum to another, the energy increases to a maximum at the transition structure and then falls. At a saddle point the first derivatives of the potential function with respect to the coordinates are all zero (just as they are at a minimum point). The number of negative eigenvalues in the Hessian matrix is used to distinguish different types of saddle point; an n th-order transition or saddle point has n negative eigenvalues. We are usually most interested in first-order saddle points, where the energy passes through a maximum for movement along the pathway that connects the two minima, but is a minimum for displacements in all other directions perpendicular to the path. This is shown schematically for a two-dimensional energy surface in Figure 5.20.

These negative eigenvalues of the Hessian matrix are often referred to as the 'imaginary' frequencies for motion of the system over the saddle point. We can illustrate this concept using the gas-phase S_N2 reaction between Cl⁻ and CH₃Cl. As the chloride ion approaches the methyl chloride along the line of the C–Cl bond the energy passes through an ion-dipole complex which is at an energy minimum. The energy then rises to a maximum at the pentagonal transition state. The energy profile is drawn in Figure 5.21. The geometries of the minimum and the pentagonal transition state, as determined by an *ab initio* HF/SCF calculation with the 6-31G* basis set are shown in Figure 5.22. The lowest-frequency eigenvalues and a representation of the corresponding eigenvectors for the two geometries are also given in Figure 5.22. There are three frequencies in the ion-dipole minimum that are of particularly low energy; two of these correspond to degenerate 'wagging' motions of the

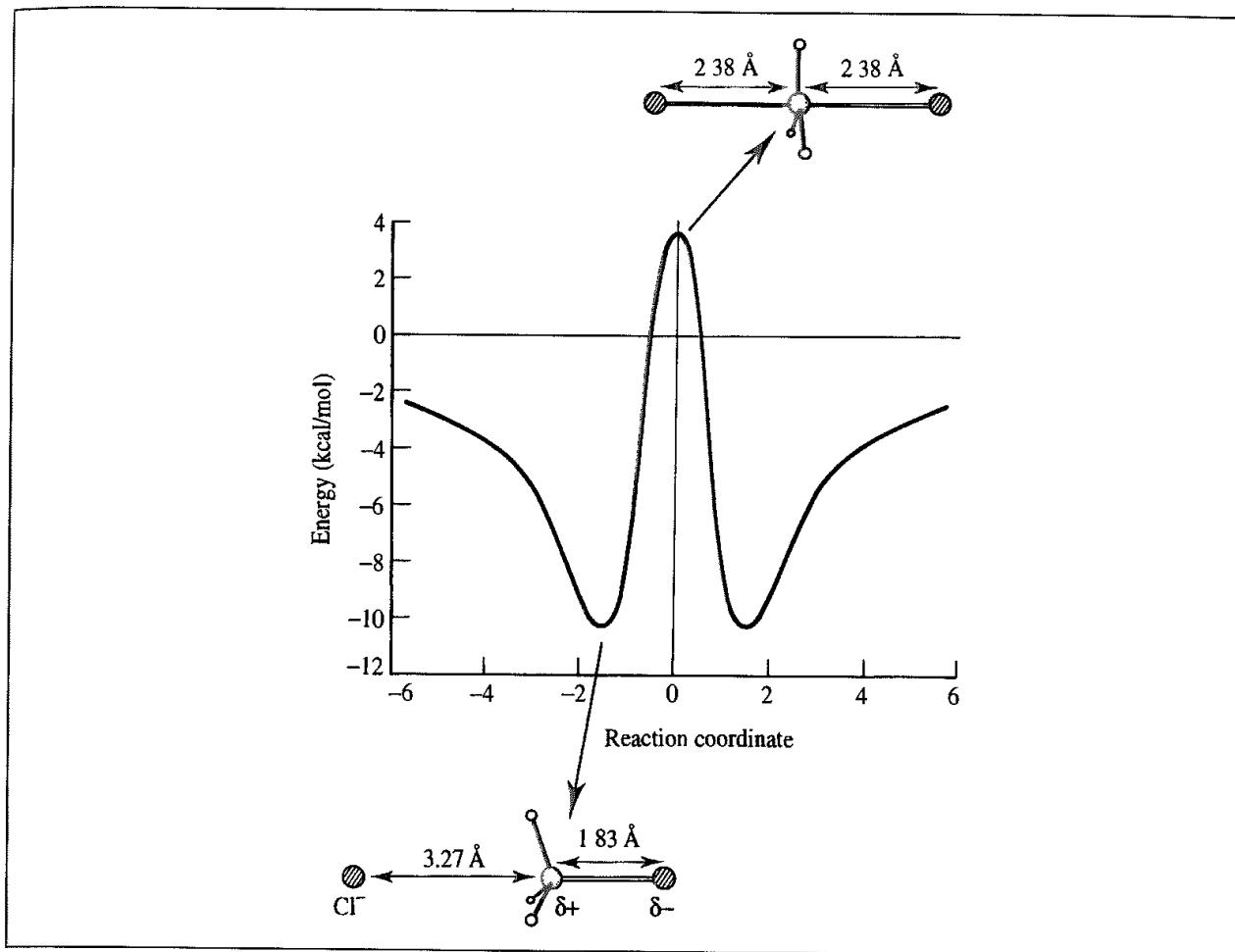


Fig 5.21. The energy profile for the gas-phase $\text{Cl}^- + \text{MeCl}$ reaction (Adapted in part from Chandrasekhar J, S F Smith and W L Jorgensen 1985 Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution Journal of the American Chemical Society 107 154–163.)

system (at 71.3 cm^{-1}). The vibration at 101.0 cm^{-1} is the normal mode that corresponds to motion towards the transition state. At the saddle point there is a single negative eigenvalue (with an imaginary ‘frequency’ of -415.0 cm^{-1}) that corresponds to vibration along the $\text{Cl}-\text{C}-\text{Cl}$ axis (i.e. motion along the reaction pathway). The other normal modes at the saddle point all have positive frequencies; the two lowest (at 204.2 cm^{-1}) correspond to wagging motions perpendicular to the $\text{Cl}-\text{C}-\text{Cl}$ axis and the third is a symmetric stretch of the two chlorine atoms along the symmetry axis.

It is important to distinguish the transition *structure* from the transition *state*. The transition structure is the point of highest potential energy along the pathway. By contrast, the transition state is the geometry at the peak in the free energy profile. In many cases the geometry at the transition state is very similar to that of the transition structure. However, the transition state may be different as the free energy of activation includes contributions from sources other than just the potential energy. If the transition state geometry is temperature-dependent then entropic factors may be important. An example is the following radical

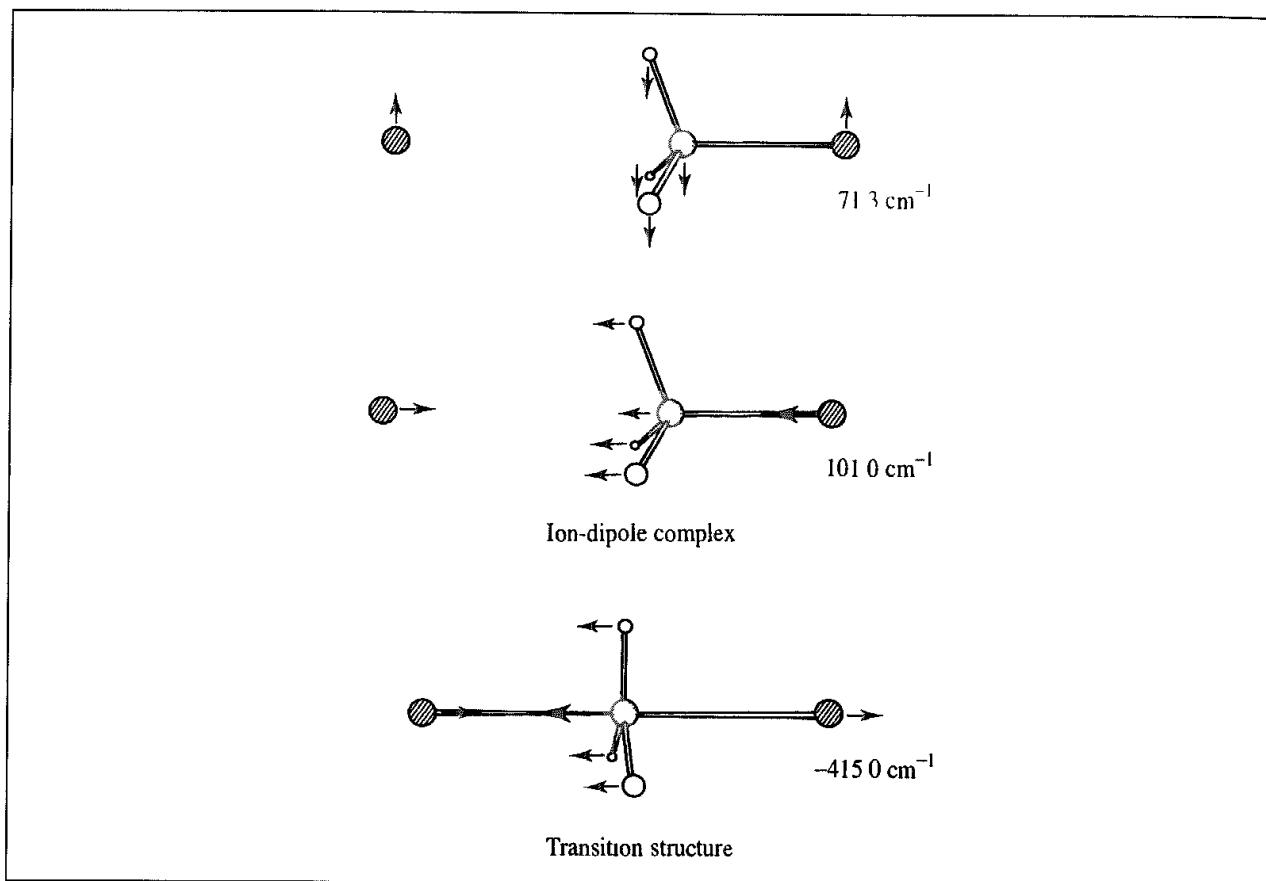
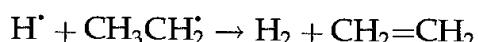


Fig 5.22 Schematic representation of some of the lower frequencies in the ion-dipole complex for the $\text{Cl}^- + \text{MeCl}$ reaction and the imaginary frequency of the transition structure, calculated using a 6-31G* basis set.

reaction:



The calculated geometry of the transition structure resembles the ethyl radical (Figure 5.23) [Doubleday *et al.* 1985]. The entropy change for this reaction is negative and so, as the temperature is increased, the maximum in the free energy profile shifts more towards the products, in the direction of lower entropy.

Methods for finding transition structures and reaction pathways are often closely related. Thus, some methods for finding the reaction pathway start from the transition structure and move down towards a minimum. Such methods must be supplied with the transition structure geometry as the starting point. Conversely, some methods for locating transition structures do so by searching along the reaction pathway, or an approximation to it. Yet other methods require neither the transition structure nor the pathway, but can determine both simultaneously from the two minima. In general, it is more difficult to locate transition structures and determine reaction pathways than to find minimum points. It is therefore crucial to check that the Hessian matrix at any proposed saddle point has the required single negative eigenvalue. Methods for locating saddle points are usually most effective when given as input a geometry that is as close as possible to the transition structure. It

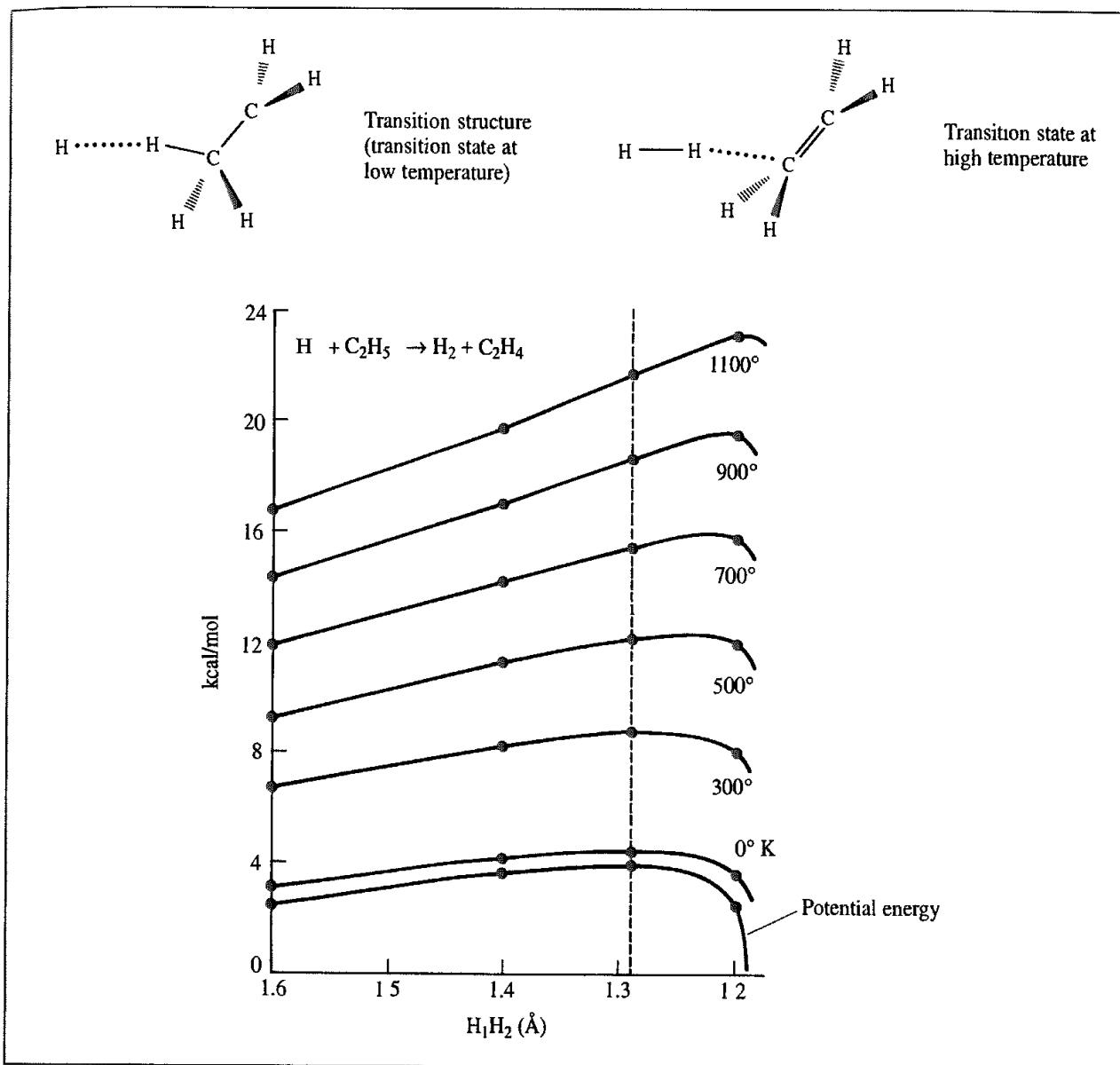


Fig 5.23 The transition structure for the $H\cdot + CH_3CH_2\cdot \rightarrow H_2 + CH_2=CH_2$ reaction. At low temperature the transition structure corresponds to the transition state (maximum of free energy). At high temperature the transition state moves closer to the products, as can be seen from the graph. (Redrawn from Doubleday C, J McIver, M Page and T Zielinski 1985. Temperature Dependence of the Transition-State Structure for the Disproportionation of Hydrogen Atom with Ethyl Radical. Journal of the American Chemical Society 107:5800–5801)

can also be helpful to examine the atomic displacements that correspond to the negative eigenvector, to ensure that it corresponds to the correct motion over the saddle point as for the $Cl^- + CH_3Cl$ reaction.

As one approaches the saddle point from a minimum, the Hessian matrix will change from having all positive eigenvalues to including one negative value. The *quadratic region* of a saddle point is that portion of the energy surface surrounding the point where the Hessian contains one negative eigenvalue. Similarly the quadratic region of a minimum is the

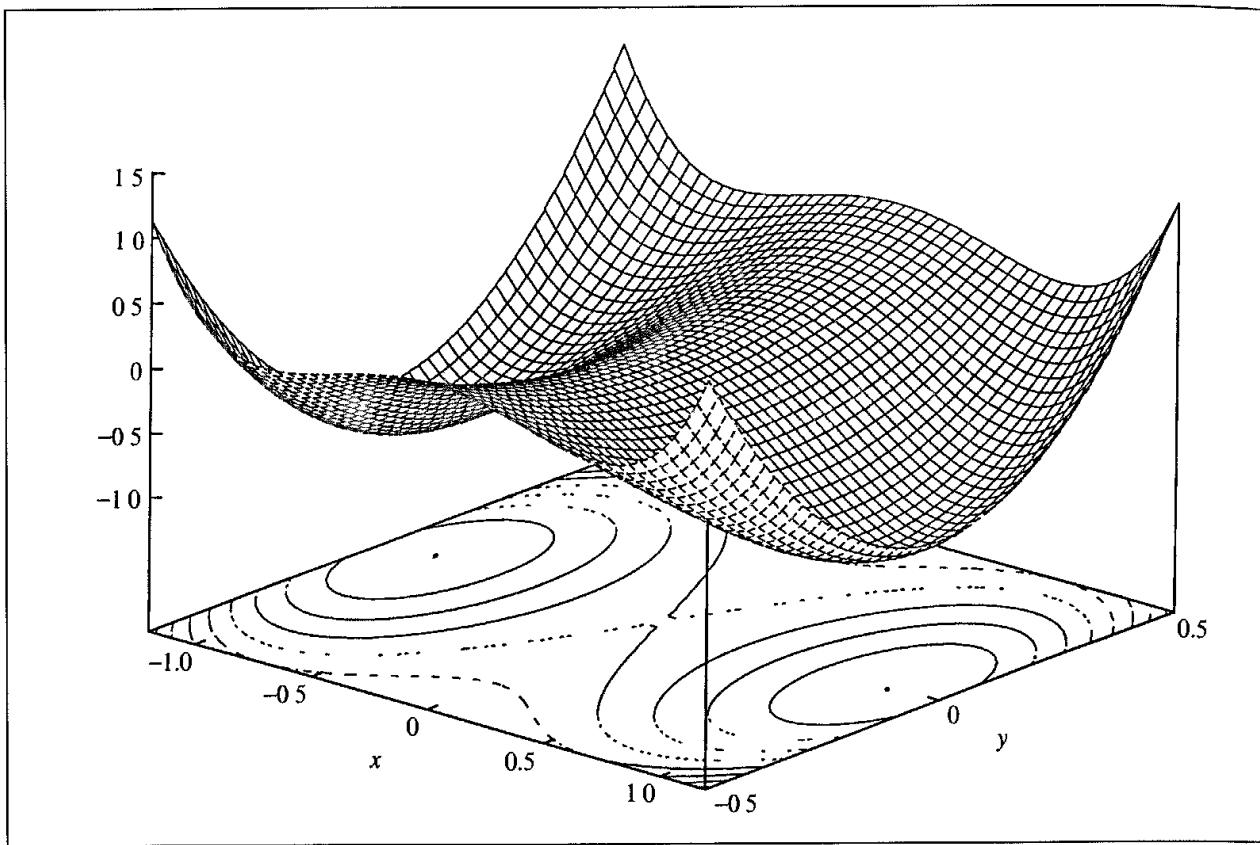


Fig. 5.24. The function $f(x,y) = x^4 + 4x^2y^2 - 2x^2 + 2y^2$ has a saddle point at $(0,0)$ and minima at $(1,0)$ and $(-1,0)$

region where all eigenvalues are positive and the Hessian is positive definite. Some algorithms for finding saddle points require a starting geometry within the quadratic region. We can illustrate the concept of a quadratic region by considering the function $f(x,y) = x^4 + 4x^2y^2 - 2x^2 + 2y^2$, which is drawn in Figure 5.24. This function has two minima at $(1,0)$ and $(-1,0)$ and one saddle point at $(0,0)$. In this case it is possible to derive and characterise the stationary points analytically. The Hessian matrix of second derivatives for this function is:

$$\begin{pmatrix} 12x^2 + 8y^2 - 4 & 16xy \\ 16xy & 8x^2 + 4 \end{pmatrix} \quad (5.39)$$

At the point $(1,0)$ the Hessian matrix is thus

$$\begin{pmatrix} 8 & 0 \\ 0 & 4 \end{pmatrix} \quad (5.40)$$

The eigenvalues of this matrix are obtained by setting the secular determinant to zero:

$$\begin{vmatrix} 8 - \lambda & 0 \\ 0 & 4 - \lambda \end{vmatrix} = 0 \quad (5.41)$$

The eigenvalues are $\lambda = 4$ and $\lambda = 8$. Thus both eigenvalues are positive and the point is a minimum. At the point $(0, 0)$ the Hessian matrix is

$$\begin{pmatrix} -4 & 0 \\ 0 & 4 \end{pmatrix} \quad (5.42)$$

with one negative and one positive eigenvalue (-4 and $+4$). The normalised eigenvectors corresponding to these eigenvalues are $(0, 1)$ for the eigenvalue $\lambda = 4$ and $(1, 0)$ for the eigenvalue $\lambda = -4$. These eigenvectors indicate the directions in which the gradient of the function changes sign. Thus along the line $x = 0$ the function passes through a minimum, as can be seen from Figure 5.24. By contrast, if one progresses from $(-1, 0)$ to $(1, 0)$ through the origin then the function passes through a maximum. As one progresses through a transition structure the eigenvector of the negative eigenvalue corresponds to the concerted motions of the atoms that give rise to motion through the saddle point. If we move along the x axis from the minimum at $(1, 0)$ to the saddle point at the origin, both eigenvalues will be positive so long as $12x^2 + 8y^2 - 4 > 0$. Thus, so long as x is larger than $1/\sqrt{3}$ the eigenvalues of the Hessian matrix will be positive. When x becomes smaller than $1/\sqrt{3}$ there will be one negative and one positive eigenvalue. In this case the quadratic region would correspond to all points where the absolute value of x was less than $1/\sqrt{3}$.

5.9.1 Methods to Locate Saddle Points

In some simple cases such as the chloride/methyl chloride reaction the geometry of the transition structure can be predicted by inspection. In other cases a *grid search* can be used to scan the energy surface in order to locate the approximate position of the transition state. In a grid search, the coordinates are systematically varied to generate a set of structures, for each of which the energy is calculated. It may then be possible to fit an analytical expression to these points, from which the saddle point can be predicted by standard calculus methods. The grid search method is widely used for constructing potential energy surfaces but is restricted to systems with a very small number of atoms or where only a limited number of degrees of freedom are being explored such as the $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ reaction. An advantage of the grid search is that it does provide information about the energy surface away from the pathway, which can be important if one wishes to investigate the dynamics of a reaction and the interconversion of energy between different modes. The grid search method is not the method of choice for all but the smallest systems due to the number of energy evaluations that are required. In any case it does not directly provide the transition structure.

The conversion of one minimum-energy structure into another may sometimes occur primarily along just one or two coordinates. In such cases, an approximation to the reaction pathway can be obtained by gradually changing the coordinate(s), allowing the system to relax at each stage using minimisation while keeping the chosen coordinate(s) fixed. The point of highest energy on the path is an approximation to the saddle point and the structures generated during the course of the calculation can be considered to represent a sequence of points on the interconversion pathway. When such coordinate driving methods are applied to conformational changes that occur primarily via rotation about bonds, the

the Hessian matrix of second derivatives is available then the appropriate direction to take is uphill along the eigenvector of the smallest eigenvalue when all eigenvalues are positive and downhill along the eigenvector corresponding to the negative eigenvalue when within the quadratic region of the saddle point [Baker 1986].

As we have stated frequently, at a saddle point the gradient is zero (as it is for a minimum). It might therefore be imagined that a minimisation algorithm (or some variant) could be used to locate saddle points. Some minimisation algorithms can occasionally incorrectly converge to a saddle point, especially if the starting structure is close to the transition structure. A simple example is the Newton-Raphson method, which will converge to a transition structure when giving a starting position that is within the quadratic region. Other minimisation algorithms can also be modified so that they consistently locate saddle points when provided with an initial structure within the quadratic region [Schlegel 1982].

5.9.2 Reaction Path Following

The traditional way to elucidate the reaction path is to move downhill from a saddle point to the two associated minima. There may be many different paths that could be followed from the saddle point to the associated minima. The *intrinsic reaction coordinate* (IRC) is the path that would be followed by a particle moving along the steepest descents path with an infinitely small step from the transition structure down to each minimum when the system is described using mass-weighted coordinates (as in a normal mode calculation) [Fukui 1981]. The initial directions towards each minimum can be obtained directly from the eigenvector that corresponds to the imaginary frequency at the transition structure. A simple steepest descents algorithm with a reasonable step size will usually give a path that oscillates about the true minimum energy path, as illustrated in Figure 5.28. This is perfectly acceptable in a minimisation, where the objective is to locate the minimum as efficiently as possible and where we are not interested in the intermediate structures. To determine the true reaction pathway (or a better approximation to it) it is necessary to 'correct' the path taken by the steepest descents algorithm. These corrective methods are especially useful when the path is curved.

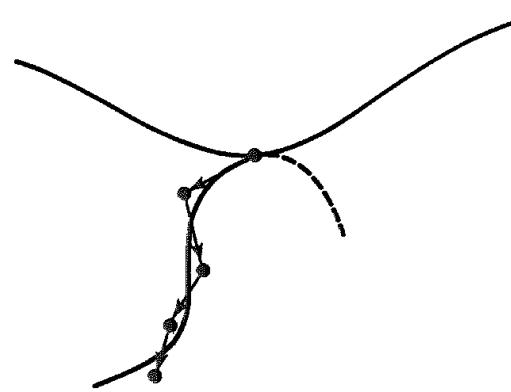


Fig 5.28 A steepest descents minimisation algorithm produces a path that oscillates about the true reaction pathway from the transition structure to a minimum

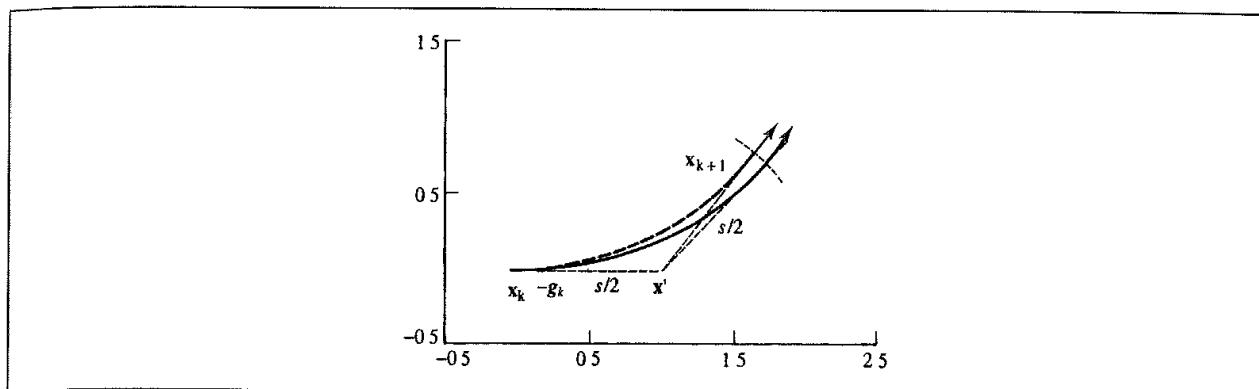


Fig 5.29: Method for correcting the path followed by a steepest descents algorithm to generate the intrinsic reaction coordinate. The solid line shows the real path and the dotted line shows the algorithmic approximation to it (Figure redrawn from Gonzalez C and H B Schlegel 1988 An Improved Algorithm for Reaction Path Following Journal of Chemical Physics **90** 2154-2161)

Many different algorithms have been suggested for determining reaction paths. The real challenge is to find an approach that is sufficiently general to work well in many (if not all) situations and with relatively little computational expense. One widely used method was devised by Gonzalez and Schlegel [Gonzalez and Schlegel 1988] and is illustrated in Figure 5.29. First it calculates the gradient at the current point, x_k . A step of length $s/2$ is taken along the direction of this gradient to give a new point (x'). The next point on the reaction path is obtained by minimising the energy subject to the constraint that the distance between x' and the new point on the reaction path (x_{k+1}) is $s/2$. The reaction path is then approximated by a circle that passes through both x_k and x_{k+1} and whose tangents at those two points are in the directions of the gradients. A refined version of this path-following algorithm has been incorporated into an efficient combined procedure which can determine reaction paths, minima and transition state geometries [Ayala and Schlegel 1997] without the need for second derivatives to be calculated.

5.9.3 Transition Structures and Reaction Pathways for Large Systems

Most of the algorithms we have discussed so far, with the possible exception of adiabatic mapping, were originally designed to be used with quantum mechanics where relatively small numbers of atoms are involved. It is often difficult to apply these methods to the study of conformational transitions. There are several reasons for this, but one important feature is that it is assumed that there is only one saddle point between the initial and final states. There may be a number of transition structures along the pathway between two conformations of a complex molecule. Here we will discuss two related methods that were originally designed to tackle this problem using molecular mechanics.

In the self-penalty walk (SPW) method of Czerminski and Elber [Czerminski and Elber 1990; Nowak *et al.* 1991] a ‘polymer’ is constructed that consists of a series of $M + 2$ ‘monomers’. Each monomer is a complete copy of the actual system and so there are $(M + 2)N$ atoms present in the calculation. The two ends of the polymer correspond to the two minima between which we are trying to elucidate the pathway (the ‘reactant’ and the ‘product’).

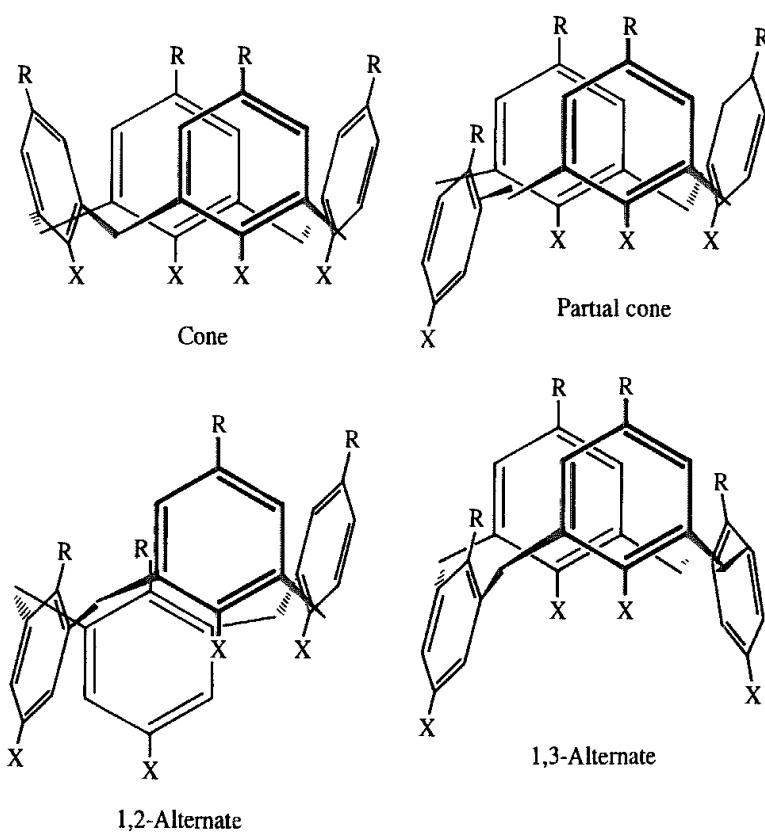


Fig. 5.31 Possible conformations of the calix[4]arene systems. (Figure adapted from Fischer S, P D J Groothuis, L C Groenen, W P van Hoorn, F C J M van Geggel, D N Reinhoudt and M Karplus 1995 Pathways for Conformational Interconversion of Calix[4]arenes Journal of the American Chemical Society 117:1611–1620)

involved, giving the energy diagram in Figure 5.32. The predicted activation barrier of 14.5 kcal/mol for the cone → inverted cone transition was in very good agreement with the experimentally determined value of 14.2 kcal/mol. Much of the barrier (9.1 kcal/mol) was due to the need to break two hydrogen bonds; the remainder was due to the need to deform some bond angles such as those of the bridging methylene carbons.

5.9.4 The Transition Structures of Pericyclic Reactions

One of the most celebrated examples of the use of quantum mechanical methods in understanding chemical reactivity is the work of Woodward and Hoffmann [Woodward and Hoffmann 1969] who were able to explain the experimentally observed nature of certain types of concerted reaction. The reactions which they studied include cycloadditions, sigma-tropic rearrangements, cheletropic reactions, electrocyclic reactions and the ene reaction (Figure 5.33) and are collectively known as pericyclic reactions. The products obtained from such reactions can be understood in terms of simple mechanistic arguments, but such arguments cannot explain some aspects. In particular, the reactions are often highly stereospecific with the reaction rates and the stereoselectivity changes dramatically with

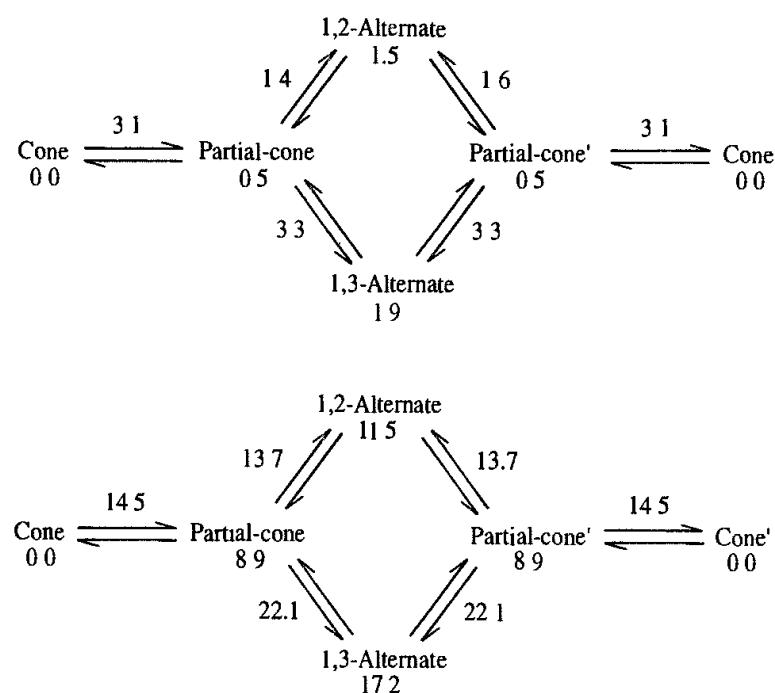


Fig 5.32 Interconversion between various conformations of calix[4]arenes $X = H, R = H$ (top); $X = OH, R = H$ (bottom). Energies in kcal/mol.

the reaction conditions. Woodward and Hoffmann successfully employed molecular orbital theory to rationalise the existing data and their theory has also been very successful in predicting the outcome of similar reactions. The basic principle applied by Woodward and Hoffmann was that of the conservation of orbital symmetry and as a consequence of their work a series of rules (often called the Woodward–Hoffmann rules) were developed. The Woodward–Hoffmann rules apply only to concerted reactions and are based upon the principle that maximum bonding is maintained throughout the course of a reaction. Fukui also discovered the importance of orbital symmetry and suggested that the majority of chemical reactions should take place at the position of, and in the direction of, maximum overlap between the highest occupied molecular orbital (HOMO) of one species and the lowest unoccupied molecular orbital (LUMO) of the other component [Fukui 1971]. These orbitals are collectively known as the *frontier orbitals*.

The HOMO–LUMO interaction depends on various factors, including the geometry of approach (which affects the amount of overlap), the phase relationship of the orbitals and their energy separation. For example, the HOMO and LUMO of ethene are illustrated pictorially in Figure 5.34. The most obvious mode of interaction between the two molecules involves suprafacial attack shown in Figure 5.34 to give cyclobutane. However, the symmetries of the overlapping orbitals must have the same phase for a favourable interaction to occur and this is not possible for ethene unless an energetically unfavourable antarafacial approach is adopted. By contrast, the interaction between ethene and the butadiene does occur in a suprafacial sense with both HOMO/LUMO pairs of orbitals having the appropriate phase relationship (Figure 5.34).

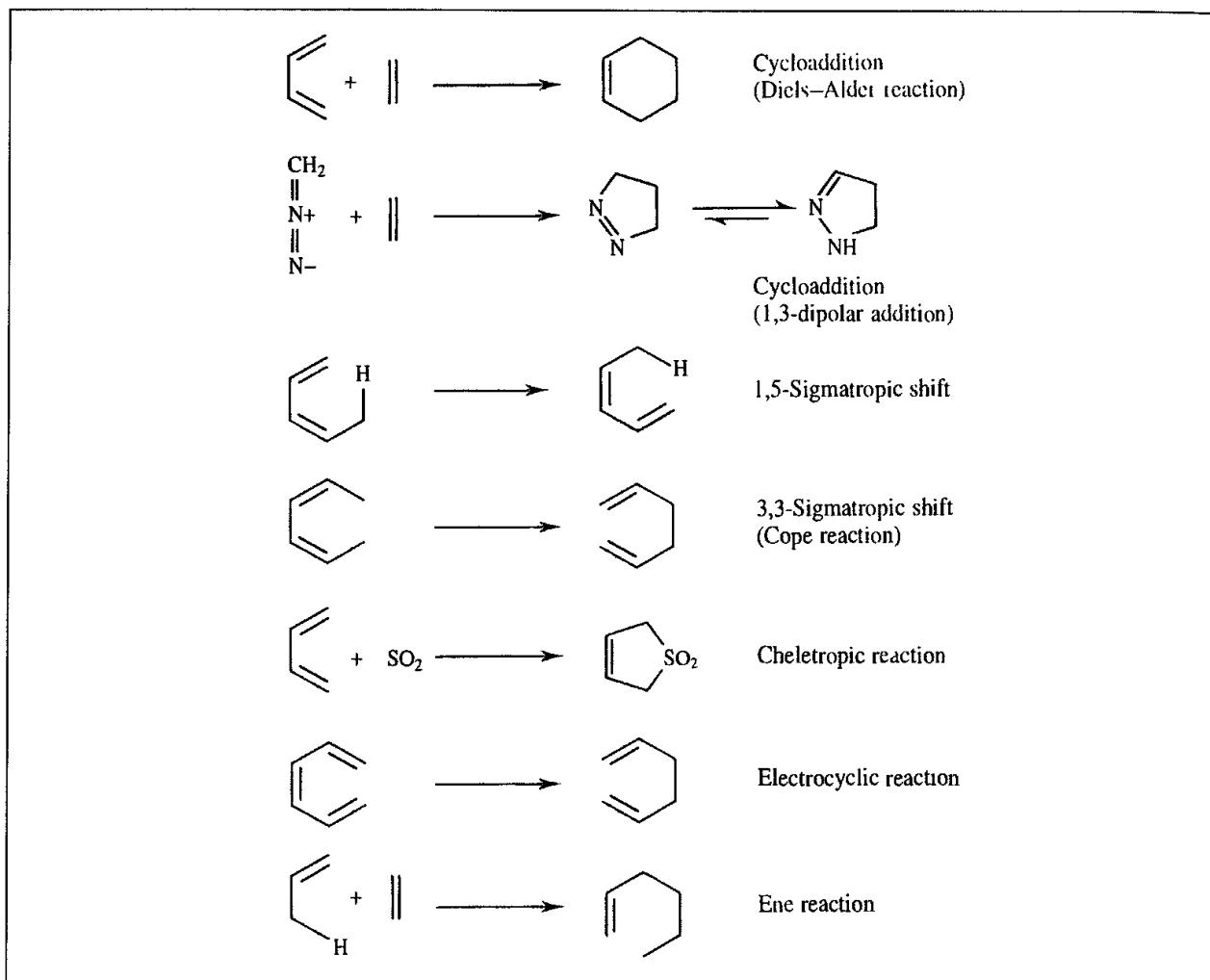


Fig. 5.33: Typical pericyclic reactions.

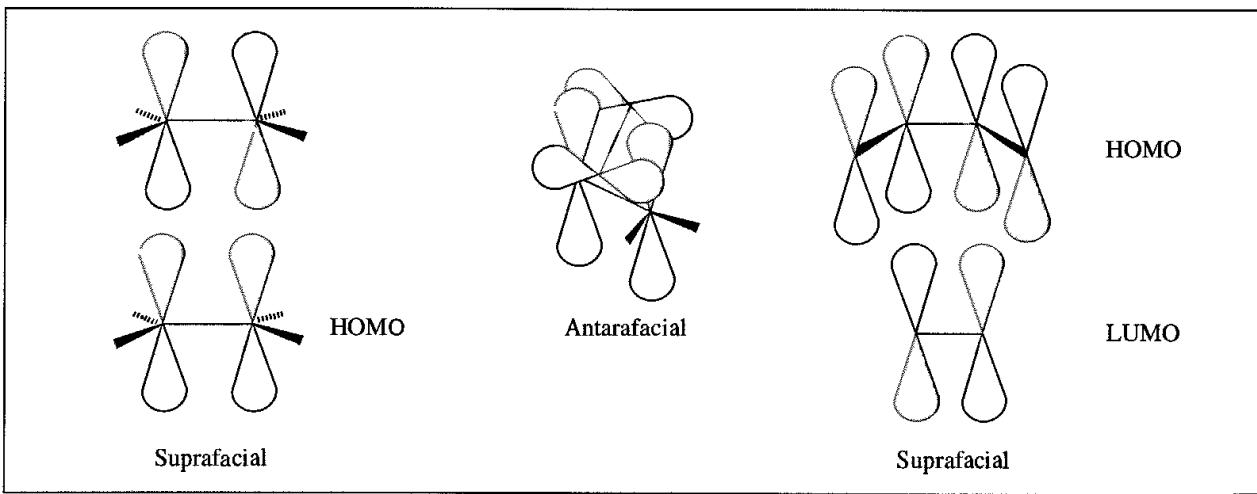


Fig. 5.34: Suprafacial attack of one ethene molecule on another (left) is not permitted by the Woodward–Hoffmann rules and the alternative antarafacial mode of attack is sterically unfavourable. Suprafacial attack is however permitted for the Diels–Alder reaction between butadiene and ethene (right).

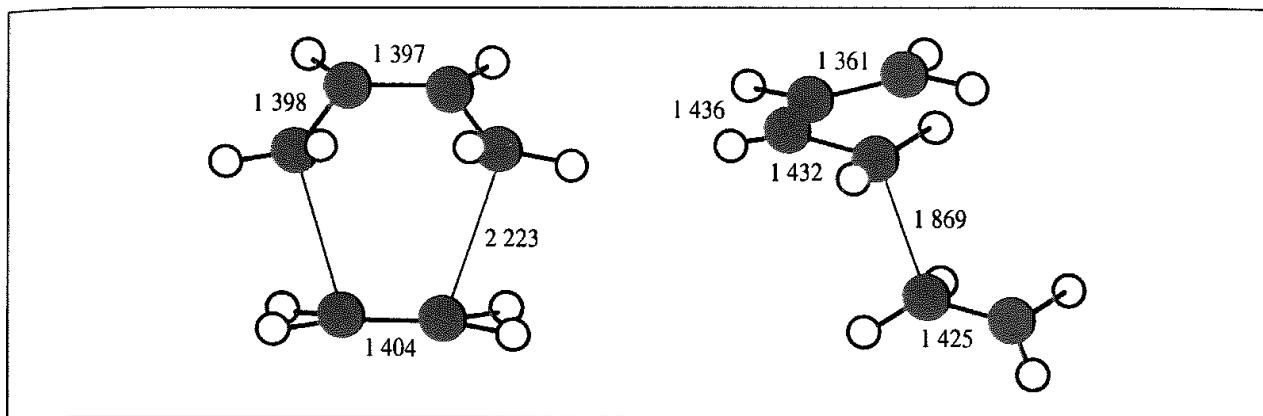


Fig 5.35 Geometry predicted by CASSCF ab initio calculations of the two possible transition structure geometries for the Diels–Alder reaction between ethene and butadiene (Figure adapted from Houk K N, J González and Y Li 1995 *Pericyclic Reaction Transition States: Passions and Punctilios 1935–1995 Accounts of Chemical Research* 28 81–90.)

The Woodward–Hoffmann rules state what the outcome of a pericyclic reaction will be, but they do not define the mechanism by which the reaction occurs. Many theoretical techniques have been applied to the study of these problems over the years [Houk *et al.* 1992] and a passionate debate has ensued on the nature of the transition structures involved in these reactions. The debate has been fuelled by the fact that different theoretical treatments (especially semi-empirical methods) give different results. For example, at one extreme the Diels–Alder reaction between butadiene and ethene would proceed via a two-step mechanism involving a biradical transition structure. At the other extreme the reaction would involve a symmetrical transition state formed in a concerted, synchronous reaction. *Ab initio* calculations at various levels of theory suggest the concerted transition structure. The geometry obtained for the prototypical Diels–Alder reaction between butadiene/ethene using a CASSCF calculation and a 6-31G* basis set is shown in Figure 5.35 [Houk *et al.* 1995]. The alternative biradical structure is also shown in Figure 5.35; this is predicted to be 6 kcal/mol higher in energy than the symmetrical transition structure.

5.10 Solid-state Systems: Lattice Statics and Lattice Dynamics

Energy minimisation and normal mode analysis have an important role to play in the study of the solid state. Algorithms similar to those discussed above are employed but an extra feature of such systems, at least when they form a perfect lattice, is that it is possible to exploit the space group symmetry of the lattice to speed up the calculations. It is also important to properly take the interactions with atoms in neighbouring cells into account.

The most straightforward type of lattice minimisation is performed at constant volume, where the dimensions of the basic unit cell do not change. A more advanced type of calculation is one performed at constant pressure, in which case there are forces on both the atoms and the unit cell as a whole. The lattice vectors are considered as additional variables along with the atomic coordinates. The laws of elasticity describe the behaviour of a material when

subjected to a *stress* (defined as the force per unit area). One obvious source of stress is any external pressure, but stress may also arise from other sources, especially from interatomic forces within the cell, which give rise to 'internal stress'. The concept of *strain* is also key to this subject; the strain is the fractional change in the dimension (for example, the change per unit length when a steel rod is stretched). In the general case we consider a situation where a point \mathbf{r} in the unstrained material moves to a new point \mathbf{r}' under the effect of some strain:

$$\mathbf{u} = \mathbf{r}' - \mathbf{r} \quad (5.48)$$

If we apply the strain uniformly in one dimension (e.g. the x axis) then the x coordinate of a point that was initially at x will change by an amount proportional to x . This is written:

$$u_x = \varepsilon_{xx}x \quad (5.49)$$

In the general case the constant of proportionality is written as the first derivative:

$$\varepsilon_{xx} = \partial u_x / \partial x \quad (5.50)$$

Deformation in the y and z directions is described in an analogous manner. In order to cater for shear-type strains additional elements are defined.

$$\varepsilon_{xy} = \varepsilon_{yx} = \frac{1}{2}(\partial u_y / \partial x + \partial u_x / \partial y) \quad (5.51)$$

These values ε give rise to a *strain tensor* (see Section 4.9.1 for more discussion on tensors), which is symmetric and is often written in the following form:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \frac{1}{2}\varepsilon_6 & \frac{1}{2}\varepsilon_5 \\ \frac{1}{2}\varepsilon_6 & \varepsilon_2 & \frac{1}{2}\varepsilon_4 \\ \frac{1}{2}\varepsilon_5 & \frac{1}{2}\varepsilon_4 & \varepsilon_3 \end{bmatrix} \quad \begin{aligned} \varepsilon_1 &\equiv \varepsilon_{xx}, & \varepsilon_2 &\equiv \varepsilon_{yy}, & \varepsilon_3 &\equiv \varepsilon_{zz} \\ \varepsilon_4 &\equiv \varepsilon_{yz}, & \varepsilon_5 &\equiv \varepsilon_{xz}, & \varepsilon_6 &\equiv \varepsilon_{xy} \end{aligned} \quad (5.52)$$

There are thus six different numbers present in the strain tensor. The symmetric form of the strain tensor prevents rotation of the unit cell with respect to the Cartesian axis system. It is possible to use this matrix to relate how a vector \mathbf{r} in the unstrained matrix is related to one \mathbf{r}' in the strained structure as follows:

$$\mathbf{r}' = (\mathbf{I} + \boldsymbol{\varepsilon})\mathbf{r} \quad (5.53)$$

\mathbf{I} is the identity matrix. The six first derivatives of the energy with respect to the strain components ε_i measure the forces acting on the unit cell. When combined with the atomic coordinates we get a matrix with $3N + 6$ dimensions. At a minimum not only should there be no force on any of the atoms but the forces on the unit cell should also be zero. Application of a standard iterative minimisation procedure such as the Davidon-Fletcher-Powell method will optimise all these degrees of freedom to give a strain-free final structure. In such procedures a reasonably accurate estimate of the initial inverse Hessian matrix is usually required to ensure that the changes in the atomic positions and in the cell dimensions are matched.

Two common properties which can be calculated from the minimum-energy structure are the elastic and dielectric constants. The elastic constant matrix is used to relate the strains of a material to the internal forces, or stresses. It is defined as the second derivative of the energy with respect to the strain, normalised by the cell volume. The inverse of the elastic

constant matrix gives the constant of proportionality between the stress and the strain. The elastic constant matrix has dimensions 6×6 and is given by the following expression:

$$\mathbf{C} = \frac{1}{V} [\boldsymbol{\gamma}_{\varepsilon\varepsilon}'' - (\boldsymbol{\gamma}_{\varepsilon r}'' \cdot \boldsymbol{\gamma}_{rr}''^{-1} \cdot \boldsymbol{\gamma}_{rz}'')] \quad (5.54)$$

In this equation $\boldsymbol{\gamma}_{\varepsilon\varepsilon}''$ is the 6×6 matrix of second derivatives (elements $\partial^2 \mathcal{V} / \partial \varepsilon_{ij}^2$), $\boldsymbol{\gamma}_{\varepsilon i}''$ and $\boldsymbol{\gamma}_{rz}''$ are the corresponding $3N \times 6$ and $6 \times 3N$ mixed coordinate/strain matrices, $\boldsymbol{\gamma}_{rr}''$ is the $3N \times 3N$ second-derivative coordinate matrix and V is the unit cell volume. It is the second term in Equation (5.54) that accounts for internal atomic relaxations as the cell distorts.

The strains on the lattice are equal to the stress divided by the elastic constant matrix:

$$\boldsymbol{\varepsilon} = (P_{\text{static}} + P_{\text{applied}}) \cdot \mathbf{C}^{-1} \quad (5.55)$$

Here we have expressed the stress as the sum of the (external) applied pressure P_{applied} together with a static pressure P_{static} , which arises from the internal forces acting on the unit cell.

The dielectric constant is concerned with the electrical properties of a material. The dielectric constant for a solid is a 3×3 matrix with different components according to the Cartesian axes. These elements are given by:

$$D_{ij} = \delta_{ij} + \frac{4\pi}{V} \mathbf{q}^T \cdot \boldsymbol{\gamma}_{rr}''^{-1} \cdot \mathbf{q} \quad (5.56)$$

In Equation (5.56) i and j are one of x , y or z ; δ_{ij} is the delta function (i.e. equal to one when $i \equiv j$ and zero otherwise) and \mathbf{q} is a vector containing the charges of each species. It is well known that the effect of a dielectric changes in an oscillating electric field (at high enough frequencies the permanent dipoles in the material are unable to keep up with the rapidly changing field). Thus one usually calculates two sets of dielectric constant matrices, corresponding to the low- and high-frequency regimes. If polarisation is included via a shell model (see Section 4.22.2) then both the cores and the shells are used to determine the low-frequency dielectric matrix; at high frequency only the shells are considered.

Comparison of the relative energies following a minimisation calculation can enable predictions to be made of the likely structure for a given material. In the same way that an organic molecule may be able to exist in more than one three-dimensional structure (or conformation – see Chapter 9) so a solid may (in principle) be able to adopt more than one three-dimensional arrangement of its atoms whilst still maintaining a periodic lattice structure. Silica, SiO_2 , has been the subject of considerable attention using these methods. The lowest-energy form is $\alpha\text{-SiO}_2$, or quartz. However, it can also form more open structures. A number of such microporous structures are in principle available, three being silicalite, mordenite and faujasite. In one study the energies of these structures relative to the quartz structure were found to be approximately 2.6, 4.9 and 5.1 kcal/mol, respectively [Ooms *et al.* 1988]. Indeed, the silicalite structure is the only one which can be prepared as the pure silicon oxide; the other forms usually require a high aluminium content and are more traditional zeolites. In an extension of this work two slightly different forms of the silicalite structure were simulated. The normal form at room temperature has orthorhombic

symmetry but at low temperatures this changes to monoclinic. These two forms are very closely related, differing only by the distortion of a key angle by 0.64° . Nevertheless, an energy-minimisation calculation starting from the orthorhombic structure did indeed change to the monoclinic, in agreement with the experimental data [Bell *et al.* 1990]. The orthorhombic \rightarrow monoclinic transition could only be observed using a force field which included polarisation effects (i.e. the shell model). Lattice minimisation methods can sometimes be very useful in helping to solve the structure of materials, a noteworthy example being the determination of the structure of a zeolite NU-87 [Shannon *et al.* 1991]. This synthetic material is of particular interest as a catalyst as it contains a multidimensional channel system. Multidimensional systems permit more complex catalytic reactions to occur and are also less prone to deactivation than one-dimensional systems. In this case, there are rings containing ten and twelve oxygen atoms (Figure 5.36 (colour plate section)). NU-87 also has a high silica content, which confers improved stability to heat. A number of experimental techniques were used to try to determine the structure, including electron diffraction and powder synchrotron X-ray diffraction, as a result of which an approximate structure was deduced but there remained some features in the powder diffraction spectrum that could not be accounted for. These were initially believed to be due to impurities but after energy-minimisation studies some subtle changes in the structure occurred to give a related structure that was a better match to the experimental data. A key feature of this particular minimisation was that the structure was not forced to adopt any specific symmetry but rather each atom was able to move independently of the others.

The calculation of vibrational frequencies (called *phonons*) is important to the study of the solid state. Indeed, the calculation of and study of phonons is often given a special name, *lattice dynamics*. To calculate the vibrational frequencies for a solid one follows a very similar approach to that described earlier for molecules, with the exception that when a shell model is being used* then their effect must be incorporated into the mass-weighted matrix of second derivatives (though not directly as they have no mass):

$$\boldsymbol{\mathcal{V}}'' = \boldsymbol{\mathcal{V}}''_{\text{core-core}} - \boldsymbol{\mathcal{V}}''_{\text{core-shell}} \cdot \boldsymbol{\mathcal{V}}'^{-1}_{\text{shell-shell}} \cdot \boldsymbol{\mathcal{V}}''_{\text{core-shell}} \quad (5.57)$$

Of additional importance is that the vibrational modes are dependent upon the reciprocal lattice vector \mathbf{k} . As with calculations of the electronic structure of periodic lattices these calculations are usually performed by selecting a suitable set of points from within the Brillouin zone. For periodic solids it is necessary to take this periodicity into account; the effect on the second-derivative matrix is that each element ij needs to be multiplied by the phase factor $\exp(i\mathbf{k} \cdot \mathbf{r}_{ij})$. A *phonon dispersion curve* indicates how the phonon frequencies vary over the Brillouin zone, an example being shown in Figure 5.37. The phonon density of states is the variation in the number of frequencies as a function of frequency. A purely transverse vibration is one where the displacement of the atoms is perpendicular to the direction of motion of the wave; in a purely longitudinal vibration the atomic displacements are in the direction of the wave motion. Such motions can be observed in simple systems (e.g. those that contain just one or two atoms per unit cell) but for general three-dimensional lattices most of the vibrations are a mixture of transverse and longitudinal motions, the exceptions

* The use of a shell model is generally recommended otherwise the resulting frequencies are too high.

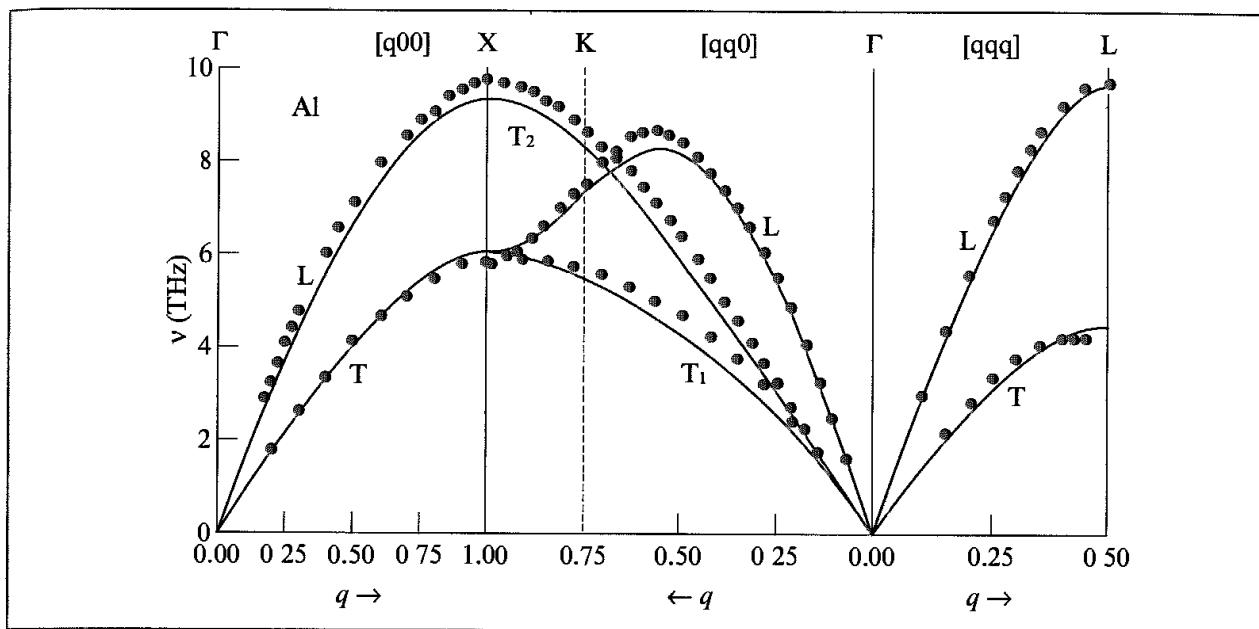


Fig 5.37. Comparison of the calculated phonon dispersion curve for Al with the experimental values measured using neutron diffraction (Figure redrawn from Michin Y, D Farkas, M J Mehl and D A Papaconstantopoulos 1999 Interatomic Potentials for Monomatomic Metals from Experimental Data and ab initio Calculations Physical Review B59 3393–3407)

being those along directions of high symmetry. The phonons are additionally classified as acoustic or optical; the former are typically of longer wavelength (lower-frequency oscillations) where the atoms move as a unit. The name arises from the fact that these are often measured as sound waves. At the point $\mathbf{k} = 0$ (the gamma point) the first three vibrational frequencies correspond to translation of the entire lattice. The optical phonons are typically higher in frequency. Various experimental techniques can be used to investigate lattice vibrations and to determine the phonon dispersion curves, the most powerful of which is inelastic scattering using thermal neutrons. These often allow the entire range of \mathbf{k} to be sampled, in contrast to some of the alternative types of radiation.

Once the phonon frequencies are known it becomes possible to determine various thermodynamic quantities using statistical mechanics (see Appendix 6.1). Here again some slight modifications are required to the standard formulae. These modifications are usually a consequence of the need to sum over the points sampled in the Brillouin zone. For example, the zero-point energy is:

$$U_{\text{vib}}(0) = \sum_{q=1}^p w_q \sum_{i=1}^{N_{\text{nm}}} \frac{\hbar\nu_i}{2} \quad (5.58)$$

In Equation (5.58) the outer summation is over the p points q which are used to sample the Brillouin zone, w_q is the fractional weight associated with each point (related to the volume of Brillouin zone space surrounding q) and ν_i are the phonon frequencies. In addition to the internal energy due to the vibrational modes it is also possible to calculate the vibrational entropy, and hence the free energy. The Helmholtz free energy at a temperature

T is thus given in the quasi-harmonic approximation by the sum of static and vibrational contributions:

$$A = \mathcal{V} + \sum_{q=1}^p w_q \sum_{i=1}^{N_{\text{hm}}} \left(\frac{\hbar\nu_i}{2} + k_B T \ln \left[1 - \exp \left(-\frac{\hbar\nu_i}{k_B T} \right) \right] \right) \quad (5.59)$$

Here, \mathcal{V} is the internal energy calculated from the potential energy model. The heat capacity at constant volume is another useful thermodynamic quantity that can be determined directly from the frequencies as it equals the derivative of the vibrational internal energy with respect to temperature.

An extension of these ideas involves minimisation of the free energy as a function of the coordinates and the temperature. The function to be minimised is sometimes referred to as an *availability*, given by $G^* = A + P_{\text{ext}} V$ where P_{ext} is the external pressure and V is the volume. Such a free energy minimisation requires derivatives of the free energy with respect to the coordinates. Early implementations used approximations such as separating the changes in the external coordinates (i.e. the dimensions of the unit cell) from the internal coordinates (i.e. the locations of the ions within the unit cell). In addition, true free energy derivatives might be calculated for the external coordinates only (due to the computational cost) with the internal coordinates being changed using the static potential energy. It is now possible to calculate a full set of analytical free energy first derivatives and hence to perform a full minimisation of the free energy with respect to external and internal coordinates simultaneously [Taylor *et al.* 1998].

Free energy minimisation provides information that is in many ways complementary to molecular dynamics simulations [Allan *et al.* 2000]. The former is particularly useful for investigating materials at lower temperatures where the harmonic assumption is valid; moreover it includes zero-point energy and quantisation effects, which are ignored by molecular dynamics. In addition, free energy minimisation provides free energies directly, rather than free energy differences, and is computationally significantly less expensive. Conversely anharmonic effects become important at higher temperatures, making molecular dynamics and Monte Carlo more suitable. One property that can be calculated via free energy minimisation is the thermal expansivity. This involves a series of free energy minimisations at different temperatures. It is also possible to calculate the free energy of disordered solids, and thus enthalpies and entropies of mixing.

Further Reading

Catlow C R A 1998. Solids: Computer Modelling. In Schleyer, P v R, N L Allinger, T Clark, J Gasteiger, P A Kollman, H F Schaefer III and P R Schreiner (Editors) *The Encyclopedia of Computational Chemistry*, John Wiley & Sons, Chichester.

Gill P E and W Murray 1981. *Practical Optimization*. London, Academic Press.

McKee M L and M Page 1993. Computing Reaction Pathways on Molecular Potential Energy Surfaces. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry Volume 4*. New York, VCH Publishers, pp. 35–65

- Press W H, B P Flannery, S A Teukolsky and W T Vetterling 1992 *Numerical Recipes in Fortran*. Cambridge, Cambridge University Press.
- Schlegel H B 1987. Optimization of Equilibrium Geometries and Transition Structures In Lawley K P (Editor) *Ab Initio Methods in Quantum Chemistry - I* New York, John Wiley & Sons, pp 249-286.
- Schlegel H B 1989 Some Practical Suggestions for Optimizing Geometries and Locating Transition States In Bertrán J and J G Csizmadia (Editors). *New Theoretical Concepts for Understanding Organic Reactions* Dordrecht, Kluwer, pp. 33-53
- Schlick T 1992 Optimization Methods in Computational Chemistry. In Lipkowitz K B and D B Boyd (Editors) *Reviews in Computational Chemistry* Volume 3. New York, VCH Publishers, pp 1-71
- Stassis C 19 Lattice Dynamics. In Sköld and D L Price (Editors) *Methods of Experimental Physics Volume 23. Neutron Scattering Part A*. Orlando, Academic Press, pp. 369-440.
- Watson G W, P Tschaufeser, A Wall, R A Jackson and S C Parker 1997. Lattice Energy and Free Energy Minimisation Techniques. *Computer Modelling in Inorganic Crystallography*. San Diego, Academic Press, pp 55-81.
- Williams I H 1993. Interplay of Theory and Experiment in the Determination of Transition-State Structures. *Chemical Society Reviews* 1:277-283.

References

- Allan N L, G D Barrera, J A Purton, C E Sims and M B Taylor 2000. Ionic Solids at High Temperatures and Pressures: *Ab initio*, Lattice Dynamics and Monte Carlo Studies. *Physical Chemistry Chemical Physics* **2**:1099-1111.
- Ayala P Y and H B Schlegel 1997 A Combined Method for Determining Reaction Paths, Minima and Transition State Geometries *Journal of Chemical Physics* **107** 375-384
- Baker J 1986. An Algorithm for the Location of Transition States. *Journal of Computational Chemistry* **7**:385-395.
- Bell R G, R A Jackson and C R A Catlow 1990. Computer Simulation of the Monoclinic Distortion in Silicalite. *Journal of the Chemical Society Chemical Communications* **10**:782-783.
- Brooks B and M Karplus 1983 Harmonic Dynamics of Proteins: Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor. *Proceedings of the National Academy of Sciences USA* **80**:6571-6575
- Czerninski R and R Elber 1990. Self-Avoiding Walk Between 2 Fixed-Points as a Tool to Calculate Reaction Paths in Large Molecular Systems. *International Journal of Quantum Chemistry* **S24**:167-186.
- Dauber-Osguthorpe P, V A Roberts, D J Osguthorpe, J Wolff, M Genest and A T Hagler 1988 Structure and Energetics of Ligand Binding to Proteins: *Escherichia coli* Dihydrofolate Reductase-Tri-methoprim, A Drug-Receptor System. *Proteins Structure, Function and Genetics* **4**:31-47
- Doubleday C, J McIver, M Page and T Zielinski 1985 Temperature Dependence of the Transition-State Structure for the Disproportionation of Hydrogen Atom with Ethyl Radical. *Journal of the American Chemical Society* **107**:5800-5801
- Elber R and M Karplus 1987. A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chemical Physics Letters* **139**:375-380.
- Fischer S, P D J Groothuis, L C Groenen, W P van Hoorn, F C J M van Geggel, D N Reinhoudt and M Karplus 1995 Pathways for Conformational Interconversion of Calix[4]arenes. *Journal of the American Chemical Society* **117**:1611-1620.
- Fischer S and M Karplus 1992 Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom. *Chemical Physics Letters* **194**:252-261.

- Fisher C L, V A Roberts and A T Hagler 1991. Influence of Environment on the Antifolate Drug Trimethoprim: Energy Minimization Studies *Biochemistry* **30**:3518–3526
- Fukui K 1971. Recognition of Stereochemical Paths by Orbital Interaction. *Accounts of Chemical Research* **4**:57–64
- Fukui K 1981. The Path of Chemical Reactions – The IRC Approach *Accounts of Chemical Research* **14**:368–375.
- Gelin B R and M Karplus 1975. Sidechain Torsional Potential and Motion of Amino Acids in Proteins: Bovine Pancreatic Trypsin Inhibitor *Proceedings of the National Academy of Sciences USA* **72** 2002–2006.
- Gonzalez C and H B Schlegel 1988. An Improved Algorithm for Reaction Path Following *Journal of Chemical Physics* **90**:2154–2161.
- Houk K N, J González and Y Li 1995 Pericyclic Reaction Transition States: Passions and Punctilios 1935–1995 *Accounts of Chemical Research* **28**:81–90
- Houk K N, Y Li and J D Evanseck 1992 Transition Structures of Hydrocarbon Pericyclic Reactions. *Angewandte Chemie International Edition in English* **31**:682–708.
- Nowak W, R Czerminski and R Elber 1991. Reaction Path Study of Ligand Diffusion in Proteins: Application of the Self Penalty Walk (SPW) Method to Calculate Reaction Coordinates for the Motion of CO through Leghemoglobin *Journal of the American Chemical Society* **113** 5627–5737.
- Ooms G, R A van Santen, C J J Den Ouden, R A Jackson and C R A Catlow 1988. Relative Stabilities of Zeolitic Aluminosilicates. *Journal of Physical Chemistry* **92** 4462–4465.
- Peng C, P Y Ayala, H B Schlegel and M J Frisch 1996. Using Redundant Internal Coordinates to Optimise Equilibrium Geometries and Transition States. *Journal of Computational Chemistry* **17**:49–56.
- Pople J A, A P Scott, M W Wong and L Radom 1993. Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6-31G* and MP2/6-31G* Harmonic Frequencies *Israel Journal of Chemistry* **33**:345–350.
- Schlegel H B 1982. Optimisation of Equilibrium Geometries and Transition Structures *Journal of Computational Chemistry* **3**:214–218
- Shannon M D, J L Casci, P A Cox and S J Andrews 1991. Structure of the Two-dimensional Medium-pore High-silica Zeolite NU-87. *Nature* **353**:417–420
- Taylor M B, G D Barrera, N L Allan, T H K Barron and W C Mackrodt 1998. Shell: A Code for Lattice Dynamics and Structure Optimisation of Ionic Crystals. *Computer Physics Communications* **109**: 135–143.
- Woodward R B and R Hoffmann 1969. The Conservation of Orbital Symmetry. *Angewandte Chemie International Edition in English* **8**:781–853.