

3 Molecular mechanics

3.1 Introduction

The calculations outlined in the previous chapter, while offering a rigorous description of the molecule from both a structural and electronic point of view, suffer as a result of their computer requirements. The amount of computer time to do conformational analysis, including minimization, or to define a potential surface even for a small molecule, can be prohibitive in terms of both time and memory. When these problems are considered in tandem with the explosion of structural data in both the biological and materials fields, it is obvious that a different representation is required. In addition, it should be borne in mind that for many applications the detailed information supplied by the quantum mechanical calculation is of little interest, especially if only a molecular geometry is required. At all times one should be pragmatic as to which method would be the best for a given application.

Just as the 1960s witnessed an explosion in the application of quantum mechanical methods, there was a complementary interest in so-called 'force field' methods for conformational analysis. The origin of these methods lies in vibrational spectroscopy, where the information derived from detailed analyses of vibrational spectra required the development of potential functions to describe the overall molecular behaviour. Two different approaches were considered. In the first, the Central Force Field (CFF) method, the molecular vibrations were fitted to a function which was a sum of pairwise interactions, without reference to the covalent structure of the molecule. The obvious disadvantage of this approach is that although such a description is correct in terms of a quantum mechanical model of a molecule, it lacks the intuitive link with structure with which chemists are more happy. The second method, the Valence Force Field (VFF), provides such a description in that the vibrational data is fitted to a potential function consisting of bond length and bond angle dependent terms. This is much more satisfactory and has the advantage of allowing comparisons between molecules; unlike the CFF potential functions which will be very molecule dependent. The major criticism of the VFF method is that the force constants produced must attempt to incorporate intramolecular interactions such as dispersion forces which result from electron correlation, and therefore are not simply a representation of the intrinsic vibrational frequency.

These spectroscopic force fields provided the ideal starting point for what is now called molecular mechanics. By bringing together features from both the CFF and VFF methods, it proved possible to derive energy functions which were at once chemically intuitive while still retaining the concept of through space attractions and repulsions.

The theoretical basis of the molecular mechanics method can be derived by taking an alternative approach to the Born-Oppenheimer approximation to that considered in molecular orbital methods: in this case the nuclear motion is considered while implying a fixed electron distribution associated with each atom. To this end a model has been developed whereby a molecule is represented as a collection of spheres (possibly deformable) joined by springs. The motions of these atoms can then be described by the laws of classical physics and simple potential energy functions can be used. This allows much larger chemical systems (of the order of thousands of atoms) to be investigated.

Although this method of calculation sounds ideal, the following caveats must be kept in mind. First, as the method neglects explicit representation of electrons, it is restricted principally to the discussion of molecular ground states. This also disallows the investigation of reactions. Secondly, the results obtained will only be as good as the potential functions and parameters used; much of the potential surface defined by the force field has little validity as, typically, only extrema (stable conformations, rotational barriers etc.) are used in the parameterization procedure.

3.2 The energy calculation

The molecular representation introduced in the previous section was one which treated molecules as a set of vibrating spheres. The next step is to define an energy function which is consistent with this concept yet allows accurate calculation of molecular properties. The force fields commonly encountered today have resulted from a number of generations of development. Typically, more structural and thermodynamic data have become available, coupled with considerable increases in computer power, allowing an extension of the functional form of the energy calculation. As more terms are included, the accuracy of the force field increases.

The energy of a molecule is calculated as a sum of the steric and non-bonded interactions present. Therefore each bond length, angle and dihedral is treated individually while non-bonded interactions represent the influence of non-covalent forces.

$$E_{\text{tot}} = E_1 + E_{\theta} + E_{\omega} + E_{\text{nb}} \quad (3.1)$$

Here E_1 , E_{θ} , E_{ω} , and E_{nb} are respectively the total bond, angle, dihedral and non-bonded energies. This is shown pictorially in Fig. 3.1.

Bond stretch

The typical vibrational behaviour of a bond is near harmonic close to its equilibrium value but shows dissociation at longer bond lengths (Fig. 3.2). The most accurate description is the Morse function

$$E_1 = \sum D_e [1 - \exp\{-\alpha(l - l_0)\}]^2 \quad (3.2)$$

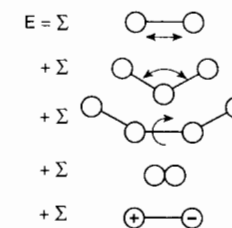


Fig. 3.1 Pictorial representation of the terms included in a molecular mechanics force field.

where l_0 is the equilibrium bond length, D_e the dissociation energy, and α a force constant. However, the exponential calculation is computationally expensive therefore most force fields have adopted a simple harmonic function

$$E_l = \sum k_l (l - l_0)^2 \quad (3.3)$$

k_l being the stretching force constant describing the deformation. The bond stretch is treated in the same fashion as a stretched spring. This equation has obvious limitations in that it only approximately describes the actual behaviour of the bond. Further, at extended bond lengths it is much too steep (see Fig. 3.2), while it provides no representation of dissociation at very large deformations. When discussing minimization of poor geometries we will see that this can be an advantage as this function can allow more extended bonds to remain intact.

Other variations on Eqn. 3.3 have been used to accommodate more accurate long distance behaviour. Most commonly this takes the form of an additional cubic term

$$E_l = \sum k_l (l - l_0)^2 + k_l' (l - l_0)^3 \quad (3.4)$$

but this suffers from the problem of inversion at long distances. Attempts have been made to remedy this by adding a quartic term which reverses the inversion.

Bond angles

Historically, bond angles have been treated in the same way as bond lengths and are usually described by a harmonic function.

$$E_\theta = \sum k_\theta (\theta - \theta_0)^2 \quad (3.5)$$

As before, k_θ is a force constant and θ_0 the equilibrium value for the bond angle. Again, this term is not ideal for the full range of values observed so higher order terms must be added. In very strained ring systems, however, it is usually not possible to use the constants derived for unstrained and acyclic molecules so separate three- and four-membered ring constants have been developed.

Dihedral angles

In very early force fields it was thought that this term could be omitted; gauche-trans energy differences would then result from non-bonded interactions. This soon proved to be an impossible task and dihedral angle terms were explicitly included. The functional form of this term is a Fourier series

$$E_\omega = \sum V_n (1 + s \cos n\omega) \quad (3.6)$$

where V_n is the rotational barrier height, n the periodicity of rotation (e.g. in ethane $n = 3$; in ethene $n = 2$) and $s = 1$ for staggered minima and -1 for eclipsed minima. Fig. 3.3 shows the $n = 1, 2$ and 3 curves.

In the simple molecules above, a single term, summed over all interactions, would suffice, but as the symmetry across the rotatable bond breaks down, the complexity of the energy profile increases. These can be corrected by the inclusion of other Fourier terms. Consider butane as 1,2-dimethyl ethane. Obviously the eclipsing interaction of the two methyl substituents will be higher in energy than that between one methyl and one hydrogen, or between two hydrogens. Increasing the size of the C-C-C three-fold barrier (V_3) would artificially modify the energy of the other eclipsing interactions. Modifying the interaction by the inclusion of a one-fold term (since the methyl-methyl eclipse occurs only once per 360° rotation) is the only option available. By way of warning, it should also be noted that the V_n parameters do not represent the complete rotational barrier but that van der Waals interactions must also be taken into account.

Non-bonded interactions

The interactions discussed in the previous sections can also be grouped together as the bonded interactions, in the sense that they are defined by the connectivity of the molecule. The non-bonded interactions, on the other hand, are distance-dependent and are calculated as the sum over all atoms with a 1,4 or greater separation. It is usual to consider these interactions as having two components: van der Waals and electrostatic. The former can be considered as both a size parameter and representative of electron correlation (resulting from instantaneous dipole interactions), while the latter provides a quantitative measurement of the influence of polarity on the energy and structure.

Many different functional forms have been used for van der Waals interactions but the most common is the so-called 6-12, or Lennard-Jones potential

$$E_{vdw} = \sum \epsilon [(r_m/r)^{12} - 2(r_m/r)^6] \quad (3.7)$$

ϵ is the well depth and r_m is the minimum energy interaction distance (Fig. 3.4). Short range repulsions are accounted for by the r^{-12} term whereas London dispersion-attraction forces are mediated by the r^{-6} component. At short distances the repulsive term dominates. The theoretical validity of this function is discussed in most physical chemistry textbooks.

Other forms have been proposed for the van der Waals interaction, principally because the r^{-12} term can be too steep at just less than optimal distances; these short contacts can be important when investigating sterically crowded structures. In the Buckingham potential

$$E_{vdw} = A \exp(-Br) - Cr^{-6} \quad (3.8)$$

an exponential replaces the repulsive r^{-12} term. In most circumstances this function behaves similarly to the Lennard-Jones equation but at very short

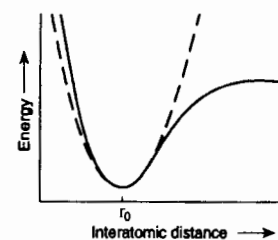


Fig. 3.2 Curves showing the variation of bond stretch energy with distance: — Morse potential; -- harmonic potential.

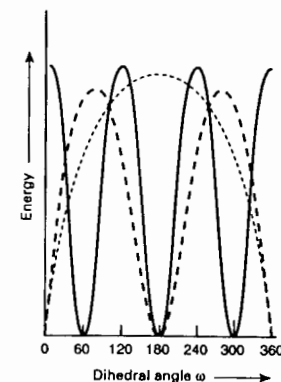


Fig. 3.3 Variation of energy with dihedral angle for one- (—), two- (---) and threefold (-·-) barriers.

interatomic distances the function inverts and goes to $-\infty$, an obvious danger in poorly constructed model structures.

The choice of function has tended to be driven by computing requirements. For a small molecule the number of interactions is relatively small and the close range behaviour is crucial. In this situation the overhead involved in calculating r , as opposed to r^2 , and the exponential is not that high. For a protein the number of interactions is considerably higher, but the contacts are likely to be close to equilibrium values. Using the Lennard-Jones function avoids the calculation of large numbers of square roots and exponentials (r^{-6} can be calculated from r^2). The 6-12 function also has the added advantage of requiring fewer parameters.

The second component of the non-bonded potential is the electrostatic term. This is usually calculated using partial charges (q) on the atom centres with the energy calculated using Coulomb's law

$$E_{el} = \sum q_i q_j / D r_{ij} \quad (3.9)$$

with the dielectric constant D taking a value appropriate to a given solvent or made proportional to the distance r_{ij} between the charges. The electrostatic contribution is one of the most controversial in molecular mechanics and will be discussed further as part of the parameterization procedure.

Other terms

The five terms outlined above constitute the core of almost all molecular mechanics force fields; in some case the entire energy function. In many situations, however, it is necessary for additional terms to be included.

For systems where hydrogen bonding is vital for stability, e.g. biological molecules, it has been common to include an additional, explicit hydrogen bond energy function to ensure correct geometries. In certain protein force fields this takes the form

$$E_{hb} = \sum (C_{ij} / r_{ij}^{12}) - (E_{ij} / r_{ij}^{10}) \quad (3.10)$$

Other force fields attempt to simulate hydrogen bonds using just the van der Waals and electrostatic terms without the inclusion of a special attractive potential. This latter method could well be the more valid as, in an attempt to retain optimum hydrogen bonding geometries, the explicit function might give correct configurations at the expense of creating strain elsewhere.

A second problem which can arise is restricting the planarity of isolated unsaturated centres. The four atoms in this grouping should be kept in a plane, however, the branch atom – e.g., in the case carbonyl groups, the oxygen – can be distorted. If the distortion is measured as the height of the central atom above the plane formed by the other three atoms then a simple restraining force can be used to hold the group in its correct geometry.

$$E_{opl} = \sum k_\chi \chi^2 \quad (3.11)$$

where k_χ is the force constant and χ the height above the plane. A modified version of this type of function can also be used as a chirality constraint.

Thus far, all of the potential functions have been concerned with isolated features of molecules. If any structural changes are correlated then, in this type of force field, they must result from a combination of appropriate forces. If only structural and thermodynamic data are required to be reproduced, this form of the force field (the so-called second generation) is adequate. However, to fit, in addition, vibrational frequencies, the coupling between geometric features must be explicitly included in the representation of the molecule.

If one considers the structure of butane (Fig. 3.5) it is clear that as the conformation changes from *anti* to *syn* there is a change in the C–C bond length and an opening of the C–C–C bond angles. The best way to incorporate this kind of feature into the force field is *via* a stretch-bend interaction. This has the effect of restraining distortion of the angle through compensatory bond stretches. A potential term of this type allows greater transferability of the isolated bond angle terms which, if fitted to butane data, would otherwise give poor ethane geometries.

Other commonly used cross terms include bend-bend and torsion-bend. The functional forms are as follows:

stretch-bend

$$E_{l\theta} = \sum \sum k_{l\theta} (l - l_0)(\theta - \theta_0) \quad (3.12)$$

bend-bend

$$E_{\theta\theta'} = \sum \sum k_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta'_0) \quad (3.13)$$

torsion-bend

$$E_{\theta\theta'\omega} = \sum k_{\theta\theta'\omega} (\theta - \theta_0)(\theta' - \theta'_0) \cos \omega \quad (3.14)$$

where the k terms are the force constants; l , l_0 , θ , θ_0 and ω are as before.

A force field including all of the valence, non-bonded and cross terms can be parameterized to give very close agreement with experiment for a large number of properties. Problems may occur, however, when highly polarizable groups or ions are present. In this situation the static charge distribution assumed in the electrostatic term is no longer realistic or adequate and an additional polarization potential will be required. Also, the van der Waals potential assumed strict pairwise interactions but more complex, many-body terms might also come into play. Attempts to include these effects are still at an experimental stage and no simple representation can be prescribed.

The final situation to be considered is the presence of delocalised π systems. This does not have any bearing on protein and other biological systems as the π systems are isolated from other unsaturated regions, e.g. aromatic amino-acids and DNA bases. This allows the use of large twofold barrier constants which will resist deformation out of the plane without interfering with the conformational mobility of other regions of the molecule. For many small, unsaturated molecules, and systems based upon porphyrins, this is not possible, unless one creates a new set of atom types for every instance. Here, some account must be taken of the degree of partial

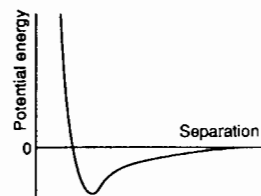


Fig. 3.4 A typical van der Waals curve.

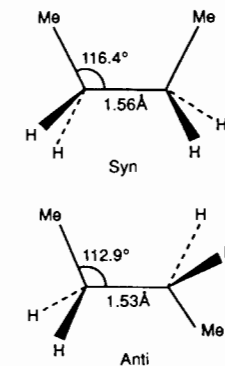


Fig. 3.5 Molecular geometries for *syn* and *anti* butane structures.

double-bond character and this should be allowed to vary according to structural changes. An elegant way of getting around this problem has been implemented in the MM2P program. The molecule in question is divided into conjugated segments and a simple SCF scheme is used to calculate bond orders about this fragment. These bond orders allow new bending, stretching, and torsional parameters to be calculated which are sensitive to the degree of local π character at any given atom. In the simplest case the parameters vary linearly with bond order. A cosine dependence can also be used for the torsion angles indicating the degree of p orbital overlap. As the structure is optimized the parameters for the π system can be recalculated if the geometry changes beyond a certain tolerance.

3.3 Energy minimization

If one has generated a model using molecular graphics, based upon standard molecular fragments, or from a z-matrix of typical internal coordinates, the energy obtained from the molecular mechanics calculation is likely to be high and not representative of the actual structure. To obtain more reliable geometries and energies one must attempt to minimize the energy of the system. This problem can be approached in two ways: either one can vary the actual internal coordinates to find their optimum value, or, as is more common, work in Cartesian coordinate space and optimize the atomic positions subject to the restraining forces generated by the molecular force field. Since most minimization methods require first, and sometimes second, derivatives of the energy the latter method is more convenient and the potential functions are easily differentiated.

A second issue which must be addressed is local versus global minimization. Given that there are three degrees of freedom per atom, for a molecule of N atoms there are $3N-6$ variables to be minimized (subtracting those due to rotation and translation). A multi-dimensional problem of this nature is further complicated by the presence of many local energy troughs on the potential surface which are minima in a mathematical sense, however, they are higher in energy than the lowest energy state, or global minimum. This situation manifests itself even in the simple rotational potential for a 1,2-di-substituted ethane (Fig. 3.6). Clearly the gauche ($\omega = 60^\circ$) conformation is stable but the molecule's preference would be for the trans conformation. This almost trivial example highlights perhaps one of the most difficult problems in computational chemistry: how does one find the global minimum (and how can one be certain that it is, in fact, the lowest energy structure).

Many of the minimization programs currently in use today are based upon the mathematical principles of the Newton-Raphson method. This requires first and second derivative information about the energy surface, but a family of algorithms exist which use different approximations to the second derivative matrix (the Hessian).

An important property of the functions used for force field calculations is that they are continuous and differentiable. From simple calculus we know

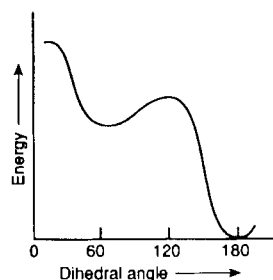


Fig. 3.6 Shape of the general rotational potential for 1,2-di-substituted ethanes.

that the condition for a minimum on a curve, the point x^* , is that the first derivative equals zero, i.e.

$$f'(x^*) = 0 \quad (3.15)$$

Since our starting point is x and not the minimum x^* we can write

$$x^* = x + \delta x \quad (3.16)$$

where δx represents the changes which x must undergo to reach the minimum value. The condition for the minimum can therefore be written in terms of x

$$f'(x + \delta x) = 0 \quad (3.17)$$

and expanded as a Taylor series

$$f'(x + \delta x) = f'(x) + f''(x)\delta x + f'''(x)\delta x^2 + \dots \quad (3.18)$$

which is also set equal to zero. Truncating the Taylor series after the second order term gives

$$f'(x) + f''(x)\delta x = 0 \quad (3.19)$$

Rearranging Eqn. 3.19 gives the expression for the change which must be made to x to reach the minimum

$$\delta x = -f'(x)/f''(x) \quad (3.20)$$

which can be substituted back into 3.16 to give

$$x^* = x - f'(x)/f''(x) \quad (3.21)$$

This is the simple one dimensional case. When considering molecules, each atom has three degrees of freedom – in x , y and z – and the term $f'(x)$ must be replaced by a $3N \times 1$ matrix (\mathbf{F}) containing terms $\delta V/\delta x_i$, the derivatives of the potential energy (V) with respect to a change in coordinate i . The corresponding second derivative matrix is constructed using all the cross-derivative terms ($\delta^2 V/\delta x_i \delta x_j$) involving each coordinate. Since matrices have replaced the single values in Eqn. 3.21, it is no longer possible to carry out the simple division. Fortunately standard computational procedures exist to derive the inverse of the Hessian matrix and the second term in Eqn. 3.21 can be replaced by $\mathbf{H}^{-1}\mathbf{F}$. Here \mathbf{H}^{-1} represents the inverse of the Hessian.

It should be noted that by truncating the Taylor series the assumption is made that the minimum is exactly quadratic in behaviour. For a complex surface this will not hold true far from the minimum but will be a better approximation as it moves closer. This forces the calculation to be carried out in a stepwise, iterative fashion, rather than reaching the minimum first time.

When working with small molecules (<100 atoms) this approach is efficient and will converge after relatively few steps. As the number of atoms is increased, the number of matrix elements in the Hessian goes up rapidly

making the calculation much slower. Additional problems with computer memory requirements for the storage of the Hessian may also be a severe limitation. This has necessitated the application of less efficient, but more practical algorithms, usually through an approximation of the Hessian. These include neglecting off-diagonal interactions between atoms (block diagonal Newton-Raphson), and the diagonal Newton-Raphson method which only calculates $\delta^2 V/\delta x^2$ values, neglecting correlation between the three degrees of freedom for a given atom.

A much more severe approximation to the Hessian is to consider it as a constant. This produces the steepest descent method which is driven purely by force gradients along the potential surface. If one imagines the energy surface as being rather like a hilly landscape then the most reliable way to find a valley, or energy well, is to follow the gradient downhill. As the gradient method has no information about the local curvature of the energy surface, minimization by this method slows down considerably as the gradient decreases. Close to the bottom of the potential well the energy differences can be rather small, however, the forces acting on the molecule can still be relatively large compared to those obtained by a less approximate method. The principal advantage of the steepest descent method is that it is very efficient when very large forces are present, far from a minimum, making it a robust choice to tidy up model geometries prior to further refinement by another method.

It is possible to exert more control in first derivative methods by including some kind of history of the minimization path. In pattern search methods the previous step is used to accelerate movement if the gradient is in the same direction as before. If the gradient changes, the pattern is abandoned and a new one set up.

A more elegant improvement to this method is that of conjugate gradients which, like pattern search, uses information from previous steps to modify the move in the next step. Unlike pattern search, it does not abandon the history if the direction changes. In the first step, where the gradient vector is \mathbf{g}_1 the move is given by

$$\mathbf{s}_1 = -\mathbf{g}_1 \quad (3.22)$$

The new direction from this step takes into account the previous gradient and follows the search direction

$$\mathbf{s}_k = -\mathbf{g}_k + \mathbf{b}_k \mathbf{s}_{k-1} \quad (3.23)$$

where \mathbf{s}_{k-1} is the search direction from the previous step and \mathbf{b}_k is a scaling factor given by

$$\mathbf{b}_k = \mathbf{g}_{i+1} \cdot \mathbf{g}_{i+1} / \mathbf{g}_i \cdot \mathbf{g}_i$$

The effect of the second term in Eqn. 3.23 is to release the constraint that the $i+1^{\text{th}}$ gradient should be orthogonal to the i^{th} gradient. Hence, better search

directions are usually obtained by this method than by steepest descents and it has much better convergence properties (Fig. 3.7).

In addition to its speed and memory requirements, the conjugate gradients method has the additional advantage that it is unlikely to maximize rather than minimize. This is one of the shortcomings of the Newton-Raphson family of methods which use only derivative information to search for stationary points: this could be either a maximum or a minimum. If a maximum on the energy surface is close to the starting point it is likely that it will be found, and a transition state structure will be obtained rather than a stable ground state.

In conclusion, the choice of minimizer should be tempered by the state of the starting structure: if it is a model, far from a minimum, then a method such as steepest descents should be applied. Once a low energy has been obtained it is advisable to switch to a method with better convergence properties – some variant of the Newton-Raphson method or conjugate gradients – which will have greater success in actually finding the minimum. Applying Newton-Raphson methods to poor structures could lead to catastrophic results, including maximized structures.

3.4 Force field parameterization

It is obvious that the quality of a given force field depends crucially on the parameters which are derived as constants in the potential function. Just as a given application can dictate which functional form is the most appropriate, the parameterization procedure must include this data when developing force constants and equilibrium values, in order to be valid. This can be both a strength and weakness of the force field method; the predictive ability of the method cannot easily extend beyond certain predetermined bounds, while within these limits the results can be very good indeed.

When discussing the different terms incorporated into the energy function, force fields were referred to as second or third generation. This reflects both the functional form of the potential equation and the scope of the data used in the fitting procedure. Available computer power must be considered as a vital factor in this historical division; obviously some kind of balance must be met between accuracy of calculation and time taken. Another consideration is the types of application of the force field. In the 1970s the principal interest was structure, which can be quite well described by the simple potential forms. More recently, interest has shifted to the dynamic behaviour of molecules. No longer is our interest restricted to the minima on the energy surface and this requires better defined descriptions between the extrema. Only by using complex energy functions which incorporate correlation between internal coordinates has this improvement been achieved.

A parameterization can be approached from two directions. One can attempt to automate the procedure and use least-squares optimization methods to obtain a simultaneous best fit of calculated results to experimental data. Like all multi-dimensional problems, local minimization is a hazard leading to mathematical solutions which have little physical

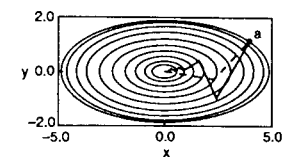


Fig. 3.7 Minimization paths for a simple energy surface: — steepest descent; -- conjugate gradients.

The development of the MM3 force field, a third generation potential, is discussed in Allinger, N. L., Yuh, Y. H., and Lii, J.-H. (1989). *J. Amer. Chem. Soc.*, **111**, 8551–8566.

significance. Optimization in this way must be closely watched to avoid excessive bias towards some parameters at the expense of others.

Alternatively, a trial and error procedure can be used. Here the user makes small changes to the parameters in an attempt to achieve the best possible fit. One advantage of this constant intervention is that a greater feel is obtained for the inter-relationships between parameters. In combination with a limited least-squares optimization better results should be possible than by the completely automated method.

As stated earlier, the data to be fitted should reflect the kind of applications for which the program will be used. For small molecules the most reliable source of data on geometries is gas phase structural studies, usually microwave or electron diffraction. The use of solution phase or crystal geometries is to be avoided as the influence of the environment is an unknown factor. Spectroscopic data can also provide information on rotational barriers and vibrational frequencies. Thermodynamic data such as heats of formation can also be included, but group contributions to ΔH_f must also be derived. Data from *ab initio* molecular orbital calculations can also provide target values if few experimental data are available. More recently, the latter method has been extended. *Ab initio* molecular energy surfaces are calculated for typical molecules using basis sets of reliable quality. Rather than using equilibrium geometries, these surfaces and their first and second derivatives have been evaluated for a collection of geometries distorted along the normal modes of vibration. These surfaces now describe not just minimum energy positions but give a more accurate representation of the complete potential energy surface. By fitting a potential energy function and its derivatives to this kind of data one can be sure that energies of non-minimum energy structures are considered as 'real' behaviour. The obvious limitation of this approach is that the surface is only as good as the *ab initio* calculations used in its derivation, but with enough computer time this ceases to be a problem. The main problem remaining is that the attractive part of a van der Waals function is not amenable to calculations at the Hartree-Fock limit.

The next issue is the origin of the starting values for each of the parameters. The important thing which must be remembered is that the resultant geometries come from the total interactions of the force field and not simply the force constant and equilibrium values for a given internal coordinate. This has the serious disadvantage that it is not strictly correct to place physical significance on the components of a force field generated energy. However, this allows greater flexibility in fitting the experimental data as the equilibrium values (l_0 , θ_0 etc.) become additional variable parameters.

A reasonable starting point for the bond-stretch and angle-bend force constants would be a consensus value from all those that are available in the spectroscopic literature; standard equilibrium values for bond lengths and angles must also be taken. For dihedral angles it should be remembered that there is an additional contribution from the non-bonded energy terms to the rotational potential. This must be subtracted before fitting V_n values.

Finding suitable values for van der Waals parameters presents one of the greatest problems in force field development. Although the close-packed geometry of a crystal would appear to provide details about interatomic distances, a simple deconvolution into atomic contributions is not easy; the close approach of hydrogen in an alkane crystal is mediated not just by H·H interactions but by C·C and C·H pairs. The resultant intermolecular arrangement is a balance over all the interactions at short range. Using heat of sublimation data, however, the crystal energetics can be calculated as a test of the parameters. Typically, however, the van der Waals radii are used as 'soft' parameters, that is to say, the exact values have less influence upon the final geometry than the high force constant bond-stretch and angle-bend terms. Therefore, these parameters can be used to fine-tune the potential.

The electrostatic interaction is unusual in that the charge density will be different for a given atom type across a range of compounds. One must either assign generic values to an atom type and make the assumption that an atom's contribution to the overall electrostatic energy is modulated by the other charges, or one can be consistent as to which method is used to calculate starting charges in the parameterization. Only charges calculated by this method will be compatible with the given force field.

Unless dealing with near ionic systems, the charges on a molecule will have less influence on the final geometry (since the electrostatic force falls off as r^{-2}) than on the energy of a system. This allows one to think in terms of molecular charge distributions rather than absolute values. In some applications (free energy calculations – Chapter 4; molecular similarity – Chapter 6) vast improvements have been found when using better quality charge distributions.

The charges themselves can be calculated by a number of methods. Empirical charge schemes, which have usually been parameterized to give close agreement with experimental dipole moment values, provide a fast way of generating charges. These are certainly suitable for structural calculations and large systems. Some force fields have been parameterized using charges obtained from molecular electrostatic potentials (see Chapter 2) and this procedure should be followed when applying the program to unknown molecules. Although there can be a tendency to use 'the best possible charges' one must always bear in mind how crucial the absolute value will be to the application. There is little point in wasting computer time on expensive *ab initio* calculations when all one wishes to do is tidy up the stereochemistry of a model structure!

A different approach to generating non-bonded parameters is to fit them to data on liquid properties; the Monte Carlo method can be used for this (Chapter 4). This method has been used to derive parameters for a wide range of molecular liquids indicating that it is a reliable procedure.

3.5 Conformational analysis

One of the most useful aspects of molecular mechanics calculations is that they allow rapid determination of molecular energies. This makes them very

For more details on using *ab initio* surfaces to generate parameters see Maple, J., Dinur, U., and Hagler, A. T. (1989). *Proc. Natl. Acad. Sci., USA*, **85**, 5350–5354.

suitable for problems of conformational analysis. Since the energies calculated are relative, in the sense that they depend entirely on the connectivity of the molecule, they cannot be compared between molecules. However, when investigating conformational space this does not arise; only internal coordinates will be changed, not connectivities.

The simplest way to carry out conformational analysis is to derive the complete energy surface which results from changing a particular rotational variable. Making the approximation that only conformations close to the maxima and minima of the rotational potential will be observed, rotations about a saturated C–C bond can be reduced to six points. For a small molecule with, say, three rotatable bonds this does not present much of a problem as only 6^3 (216) conformations will be generated. If a larger molecule, containing six rotatable bonds, is of interest the number of conformations increases dramatically to 6^6 (46,656); and this is still a very small molecule. Obviously this kind of detailed grid scan must be modified for general use. Also, if more than two variables are used it is difficult to visualise the energy surface being mapped out.

Most of the solutions to the combinatorial problem of conformational search rely on methods to filter out poor conformations at as early a stage as possible. Reducing the number of values per rotatable bond is an obvious answer but immediately it should be recognised that one runs the risk of overlooking potentially interesting structures. In the case of amino acids, each residue to be calculated can be considered a pair of interdependent variables, with only certain minima from the conformational energy map (Ramachandran map) accessible. By reducing the number of pairs to either minima, or minima within a threshold of the global minimum, one can reduce interesting conformational space to a very few points, of the order of 5–10 per residue. Similar restrictions can be applied to any pair of dihedral angles. A conformational search procedure can then proceed by considering combinations of the minima along the given chain.

A second method of increasing the efficiency of a conformational search is to use a tree-search. Here, the search starts out as a complete grid search but as individual conformations are being generated checks are carried out to eliminate unsuitable structures before completion. Typical tests would include checking for adverse non-bonded contacts. Since many large systems contain cyclic regions, ring closure is a particularly good criterion. As a conformation is being generated it should be possible to tell whether it will ever be capable of cyclizing, based upon the maximum length of a unit in the chain (Fig. 3.8).

Usually the conformations generated by any of the above methods will have rigid geometries. Hence, minimization will be required to produce more realistic structures. If only minima are required then one can proceed as in Section 3.3. If, however, one is interested in mapping out a complete surface, and barrier heights and interconversion pathways are of interest, means must be found to hold dihedrals at a particular value while minimizing all other variables. The simplest way to do this is to modify the rotational potential of the fixed dihedral by applying an offset c_2 to the minimum value and using a

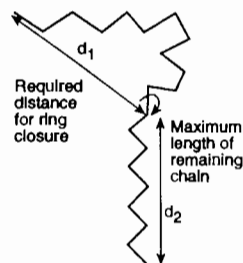


Fig. 3.8 Ring closure criteria for filtering out poor conformations during tree-searches: d_1 must be less than d_2 to continue the search.

very large rotational barrier constant c_1

$$E = c_1[1 + \cos 3(\omega + c_2)] \quad (3.25)$$

Following minimization the energy must be recalculated using the correct rotational potential. Great care must be taken when applying this method, especially when investigating cyclic molecules as the pathway across a surface may depend upon the starting point and the location of other stationary points on the energy surface.

Having obtained a potential energy surface, or conformational reaction coordinate, by one of the above methods, it is tempting to place an interpretation on the different contributions to the barrier height etc. This must be approached cautiously as it should be remembered that only features such as total energy, geometry and vibrational frequencies can be considered to have any physical meaning.

In the next chapter we will see that Monte Carlo and molecular dynamics methods, which allow molecules to traverse energy barriers, are more useful for larger systems.

3.6 Summary

Force field methods of the type outlined in this chapter are finding increasing application in many areas of structural chemistry. The principal advantage of these methods is that classically based potential energy functions are less computationally expensive than quantum mechanical methods, allowing systems of the order of thousands of atoms to be considered. A large body of data now exists validating the functional forms and the variable parameters used so that, as long as one is aware of the limitations associated with a given force field, reliable results may be obtained. Many of the older force fields will be valid only for structures while thermodynamic data can be calculated from second generation force fields such as MM2. If one wants to know more about vibrational frequencies, a force field containing cross terms in the energy function must be applied. Typically, this type of force field will be more reliable between stationary points on the energy surface.

Minimization of structures is usually required to remove strain from poorly defined geometries. When far from a minimum, simple gradient techniques, and a less complex force field expression are the methods of choice. But, as one approaches the minimum, more sophisticated methods such as conjugate gradients are recommended: these converge much more quickly. It is also important that the parameters used are consistent with the rest of the force field. Since the parameters arise from a fitting procedure they can have little physical significance, A 'mix and match' approach is to be avoided. Conformational search procedures often utilize molecular mechanics calculations but one must be careful that the methods used do not introduce bias into the calculation otherwise a true representation of the energy surface will not be obtained, or low energy conformers will be omitted.