

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

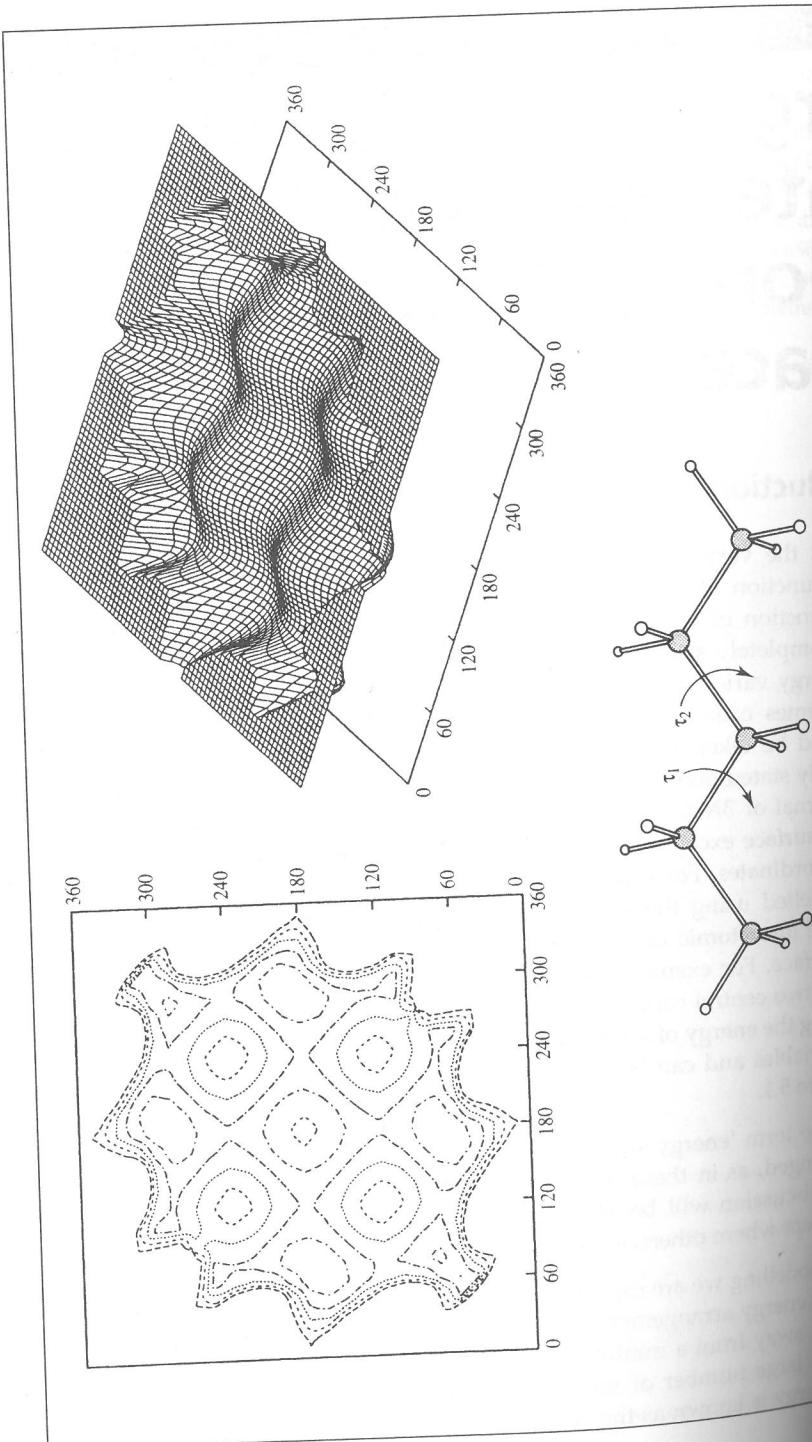


Fig. 5.1: Variation in the energy of pentane with the two torsion angles indicated and represented as a contour diagram and isometric plot. Only the lowest-energy regions are shown.

the system t
algorithm. Th
approaches interested to
For example structural ch
the pathway with the arra
points are sta
function is ze

A geographic encounter in
valleys. A mi
and featureless
taking steps '1

5.1.1 Energy

The minimisat
on one or more
has a minimum
each of the var

The functions c
energy with th
Molecular mec
where the energ
(as defined in
minimum of a
not generally p
energy varies w
which graduall
energies until t
rithms operate.
This function is
mum point, loc
from the point
of illustration, a
many more vari
We can classifi
energy with res

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

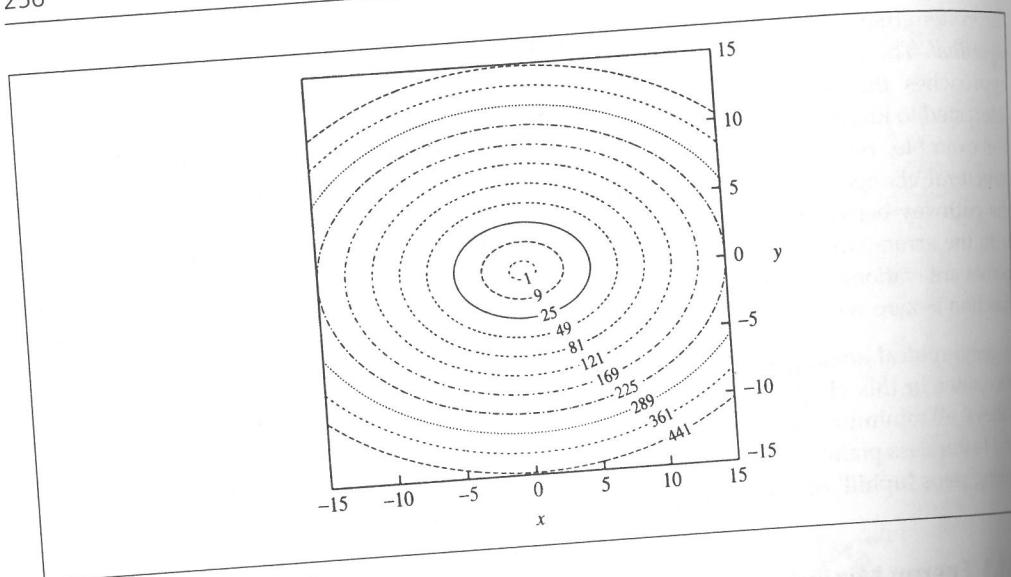


Fig. 5.2: The function $x^2 + 2y^2$.

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. The statistical weight of the global energy minimum may be important if there are many local minima.

global energy minimum may be important if there are many local minima. The global energy minimum is the one that is lower than a broad minimum. It is the minimum that is widely spaced in the energy landscape. The global energy minimum is the one that is the most stable and corresponds to the most common structure (e.g. the most common conformation).

The input to a minimisation algorithm is the initial coordinates of the atoms in the molecule. An experimental technique is to use a theoretical method to calculate the energy of the molecule at different sets of experimental and simulated coordinates. The behaviour of a protein in a solvent 'bath' can be studied by calculating its energy from a Monte Carlo simulation.

5.1.2 Derivative-based minimisation

In order to use a derivative-based minimisation algorithm, it is necessary to calculate the derivatives of the energy function with respect to the internal coordinates of the molecule. These derivatives can be calculated numerically. The numerical derivatives can be calculated using finite differences, which may be more efficient than analytical derivatives. Numerical derivatives were discussed in

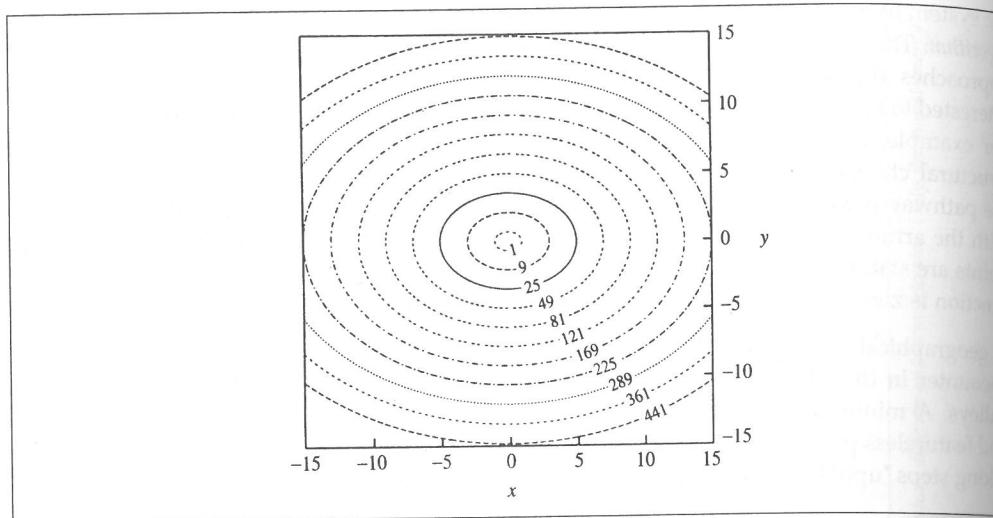


Fig. 5.2: The function $x^2 + 2y^2$.

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

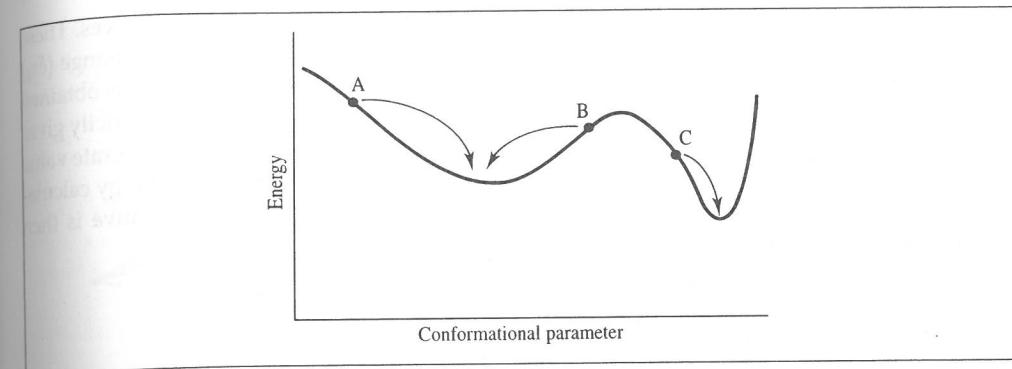


Fig. 5.3: A schematic one-dimensional energy surface. Minimisation methods move downhill to the nearest minimum. The statistical weight of the narrow, deep minimum may be less than a broad minimum which is higher in energy.

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 (δx_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 contract in one dir
S A Teukolsky and

Let us consider
(Figure 5.5). S
and (9, 11), wh
in turn. The v
vertex with th
reflected thro
(11, 7) and a

Fig. 5.5: The fin
the triangle 123
the third step is