**Lecture VI**

### Creation of new medicinal substances, main stages and ways

### Introduction

The development of powerful desk top and larger computers has enabled chemists to predict the structures and the values of properties of known, unknown, stable and unstable molecular species using mathematical equations. These equations are obtained using so called ‘models’ of the system being studied (see sections 5.2 and 5.3). Solving these equations gives the required data. The reliability of the mathematical methods used to obtain and solve the equations is well known and so in most cases it is possible to obtain a reliable estimate of the accuracy of the results. In some cases the calculated values are believed to be more accurate than the experimentally determined figures because of the higher degree of experimental error in the experimental work. Graphics packages that convert the data for the structure of a chemical species into a variety of easy to understand visual formats have also been developed (Figure 5.1). Consequently, in medicinal chemistry, it is now possible to visualize the three dimensional shapes of both the ligands and their target sites. In addition, sophisticated computational chemistry packages also allow the medicinal chemist to evaluate the interactions between a compound and its target site before synthesizing that compound (see section 5.5). This means that the medicinal chemist need only synthesize and test the most promising of the compounds, which considerably increases the chances of discovering a potent drug. It also significantly reduces the cost of development.

Molecular modelling is a complex subject and it is not possible to cover it in depth in this text. For workers wishing to use it as a tool in drug design it will be necessary to either ask a competent computational chemist to make the necessary calculations and graphic conversions or to treat the computer as a black box and use the relevant computer program according to its manufacturer’s instructions. In both approaches to molecular modelling, it is essential that the drug designer has a basic understanding of the fundamental concepts of the methods used in order to avoid making incorrect deductions, as well as to appreciate the limita- tions of the methods.

CH3

C O

O COOH

1. Stick model of aspirin. **(b)** Space fill model model of aspirin (CPK model) CH3

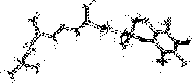
CH3

H CH3 H

O

CH3 CH3

OH CH3

**(c)** Stick model of Vitamin E. **(d)** Space fill model model of vitamin E.

Figure 5.1 Examples of some of the formats used by graphics packages to display molecular models on computer screens

#### Molecular modelling methods

#### The three dimensional shapes of both ligand and target site may be determined by X-ray crystallography or computational methods. The most common com- putational methods are based on either molecular or quantum mechanics. Both these approaches produce equations for the total energy of the structure. In these equations the positions of the atoms in the structure are represented by either Cartesian or polar coordinates (Figure 5.2). In the past, the initial values of these atomic coordinates were set by the modeller. However, as it is now customary to construct models from existing structural fragments (see section 5.2.1) modern computer programs will automatically set up the coordinates of the atoms in the first fragment from the program’s database. As additional fragments are added, the the computer automatically adjusts the coordinates of the atoms of these additional fragments to values that are relative to those of the first fragment, since it is the relative positions of the atoms that is important as regards the energy of the structure, and not the absolute positions of the atoms. Once the energy equation is established, the computer computes the set of coordinates which correspond to a minimum total energy value for the system. This set of coordinates is converted into the required visual display by the graphics package (Figure 5.1). However, although the calculations made by computers are always accurate, the calculated result should be checked for accuracy against experimental observations. In this respect it is essential that the approximations on which the calculations are based are understood. For example, most calculations are based on a frozen molecule at 0 K in a vacuum and so do not take into account that the structure is vibrating or the influence of the medium in which the chemical species is found. Calculations taking these factors into account would undoubtedly give a more realistic picture of the structure.

Quantum mechanics calculations are more expensive to carry out because they require considerable more computing power and time than molecular mechanics calculations. Consequently, molecular mechanics is the more useful source of the large structures of interest to the medicinal chemist and so this chapter will concentrate on this method. To save time and expense, structures are often built up using information obtained from databases, such as the Cambridge and Brookhaven databases. Information from databases may also be used to check the accuracy of the modelling technique. However, in all cases, the accuracy of the structures obtained will depend on the accuracy of the data used in their determination. Furthermore, it must be appreciated that the molecular models produced by computers are a caricature of reality that simply provide us with a useful picture for design and communication purposes. It is important to realize that we still do not know what molecules actually look like.

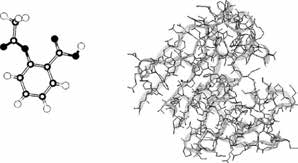
Computer graphics

In molecular modelling the data produced are converted into visual mages on a computer screen by graphics packages. These images may be displayed as space fill, CPK (Corey–Pauling–Koltun), stick, ball and stick, mesh and ribbon (see Figure 5.1 and Figure 5.3(a), 5.3(b) and 5.3(c) ). Ribbon representations are usually used to depict large molecules, such as nucleic acids and proteins. Each of these formats can, if required, use a colour code to represent the different elements, for example, carbon atoms are usually green, oxygen red and nitrogen blue. However, most graphics packages will allow the user to change this code. The program usually indicates the three dimensional nature of the molecule by making the colours of the structure lighter the further it is from the viewer. Structures may be displayed in their minimum energy or other energy states. They may be shrunk or expanded to a desired size as well as rotated about either the x or y axis. These facilities enable the molecule to be viewed from different angles and also allows the structure to be fitted to its target site (see section 5.5). In addition, it is possible using molecular dynamics (see section 5.4) to show how the shape of the structure might vary with time by visualizing the natural vibrations of the molecule (Figure 5.3(d) ) as a moving image on the screen. However, it is emphasized that both the stationary and moving images shown on the screen are useful caricatures and not pictures of the real structure of the molecule.

### Molecular mechanics

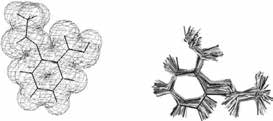
### Molecular mechanics is the more popular of the methods used to obtain molecu- lar models as it is simpler to use and requires considerably less computing time to produce a model. The molecular mechanics method is based on the assumption that the relative positions of the nuclei of the atoms forming a structure are determined by the forces of attraction and repulsion operating in that structure. It assumes that the total potential energy (*E*Total) of a molecule is given by the sum of all the energies of the attractive and repulsive forces between the atoms in the structure. These energies are calculated using a mechanical model in which these atoms are represented by balls whose mass is proportional to their relative atomic masses joined by mechanical springs corresponding to the covalent bonds in the structure. Using this model, *E*Total may be expressed mathematically by equa- tions, known as force fields. These equations normally take the general form:

**(a)**



**(b)**

**(c)**



**(d)**

Ball and stick representation of aspirin. (b) Ribbon representation of dihydrofo- late reductase. (c) Mesh representation of aspirin. This representation shown simultaneously both the space fill and stick structures of the molecule. (d) Molecular dynamics representation of aspirin at 500 K. The relative movement of the atoms with time within the molecule is indicated by the use of multiple lines between the atoms

where EStretching is the bond stretching energy (Figure 5.4), EBend is the bond energy due to changes in bonding angle (Table 5.1), ETorsion is the bond energy due to changes in the conformation of a bond (Table 5.1), EvdW is the total energy contribution due to van der Waals forces and ECoulombic the electrostatic attractive and repulsive forces operating in the molecule between atoms carrying a partial or full charge. Other energy terms, such as one for hydrogen bonding, may be added as required. Each of these energy terms includes expressions for all the specified interactions between all the atoms in the molecule.

Equilibrium bond length



*r*0

Stretched bond length *E*



*r*1



*r*

Compressed bond length

*r*1 *r*0 *r*

Bond stretching and compression related to the changes in the potential energy (*E* ) of the system

The values of each of the energy terms in Equation (5.1) are calculated by considering the mechanical or electrical nature of the structure that the energy term represents. For example, the *E*Stretching bond stretching energy for a pair of atoms joined by a single covalent bond may be estimated by considering the bond to be a mechanical spring that obeys Hooke’s law. If *r* is the stretched length of the bond and *r*0 is the ideal bond length, that is the length the bond wants to be, then:

*E*Streching ¼ 1 *k*(*r* *r*0)2 (5:2)

2

where *k* is the force constant, which may be thought of as being a measure of the strength of the spring, in other words a measure of the strength of a bond. For example, C–C bonds have a smaller *k* value than C ¼ C bonds, that is C ¼ C bonds are stronger than C–C bonds. In reality, more complex mathematical expressions, such as those given by the Morse function, would probably be used to describe bond stretching.

The value of *E*Stretching in the force field equation (see equation (5.1) ) for a structure is given by the sum of appropriate expressions for *E* for every pair of bonded atoms in the structure. For example, using the Hooke law model, for a molecule consisting of three atoms bonded a–b–c the expression would be:

*E*Stretching ¼ *E*a b þ *E*b c (5:3) that is, the expression for *E*Stretching in the force field for the molecule would be:

*E*Stretching ¼ 1 *k*(a b)(*r*(a b) *r*0(a b))2 þ 1 *k*(b c)(*r*(b c) *r*0(b c))2 (5:4)

2

2

The other energy terms in the force field equation for a structure are treated in a similar manner using expressions appropriate to the mechanical or electrical

Some of the expressions commonly used to calculate the energy terms given in equation model on which the energy term is based (Table 5.1). These expressions may be the equations given in Table 5.1 but, depending on the nature of the system beng modelled, other equations may be a more appropriate way of mathematically describing the mechanical or electrical model.

The values of the parameters *r*, *r*0, *k*, . . . etc used in the expressions for the energy terms in Equation (5.1) are either obtained/calculated from experimental observations or calculated using quantum mechanics using best fit methods. Experimental calculations are based on a wide variety of spectroscopic tech- niques, thermodynamic data measurements and crystal structure measurements for interatomic distances. Unfortunately, values are often difficult to obtain since accurate experimental data are not always available. Quantum mechanical calculations can be used when experimental information is not available but are expensive on computer time. However, this method does give better values for structures that are not in the minimum energy state. The best fit values are obtained by looking at related structures with known parameter values and using the values from the parts of these structures that most resemble the structure being modelled. Parameter values are also stored in the data bases of the molecular modelling computer programs.

#### Creating a molecular model using molecular mechanics

Molecular models are usually created by either using an existing commercial force field computer program or assembling a model from structural fragments held in the database of a molecular modelling program. In the former case commercial packages usually have several different force fields within the same package and it is necesary to pick the most appropriate one for the structure being modelled. To use the commercial force field, the values of the relevant parameters together with the initial atomic coordinates are fed into the force field equation. These values are used by the computer to calculate an initial value of ETotal for the model. This initial energy value is minimized by the computer iteratively (consecutive repetitive calculations), changing the values of the atomic coordinates in the equation for the force field until a minimum energy value is obtained. The values of the atomic coordinates corresponding to this minimum energy value are used to visualize the model on the monitor screen in an appropriate format.

The second method assembles the initial model from models of structural fragments held in the database of a molecular modelling program (Figure 5.5). Initially these fragments are put together in a reasonably sensible manner to give a structure that does not allow for steric hindrance. At this point it is necessary to check that the computer has selected atoms for the structure whose configur- ations correspond to the types of bonding required in the structure, in other words, if an atom is double bonded in the structure, the computer has selected a form of the atom that is double bonded. These checks are carried out by matching a code for the atoms on the screen against the code given in the manual for the program and replacing atoms where necessary. At this stage the structure displayed is not necessarily in its minimum potential energy conformation. However, the program can be instructed to iteratively change the atomic coordinates of the model to give a minimum value for ETotal. As a result of this change, the structure on the monitor screen assumes a conformation corresponding to a minimum energy state. This conformation may be presented in a number of formats depending on the requirements of the modeller (see Figure 5.1 and Figure 5.3(a)–(c) ).

The energy minimizing procedure also automatically twists the molecule to allow for steric hindrance. However, the energy minimizing process is not usually very sophisticated. It stops when the force field reaches the nearest local minimum energy value even though this value is not necessarily the lowest minimum energy value for the structure.

Consequently, it may be necessary to use a more sophisticated computer procedure, molecular dynamics (section 5.3), to obtain the lowest minimum energy value and as a result the best model for the molecule. This final structure may be moved around the screen and expanded or reduced in size. It can also be rotated about the *x* or *y* axis to view different elevations of the molecule.

The molecular mechanics method requires considerably less computing time than the quantum mechanical approach and may be used for large molecules containing more than a thousand atoms. This means that it may be used to model target sites as well as drug and analogue molecules. As well as being used to produce molecular models, it may also be used to provide information about the binding of molecules to receptors (see section 5.5) and the conformational changes (see section 5.3) that occur in the molecule. However, molecular mech- anics is not so useful for computing properties, such as electron density, that are related to the electron cloud. Furthermore, it is important to realize that accuracy of the structure obtained will depend on the quality and appropriate- ness of the parameters used in the force field. Moreover, molecular mechanical calculations are normally based on isolated structures at 0 K and do not nor- mally take into account the effect of the environment on the structure.

### Molecular dynamics

Molecular mechanics calculations are made at 0 K, that is on structures that are frozen in time and so do not show the natural motion of the atoms in those structures. Molecular dynamics programs allow the modeller toshow the dynamic nature of molecules by simulating the natural motion of the atoms in a structure. This motion, which is time and temperature dependent, is modelled by includ- ing terms for the kinetic energy of the atoms in the structure in the force field by using equations based on Newton’s laws of motion. The solution of the these force field equations gives coordinates that show how the positions of the atoms in the structure vary with time. These variations are displayed on the monitor in as a moving picture. The appearance of the this picture will depend on the force field selected for the structure and the temperature and time interval used for the integration of the Newtonian equations. Molecular dynamics can also be used to find minimum energy structures (Figure 5.6) and conformational analysis.

#### Conformational analysis

Each frame of the molecular dynamics ‘movie’ corresponds to a conformation of the molecule, which may be displayed on the monitor screen in any of the set formats. The program is also able to compute the total energy of each of these conformations and plot a graph of energy against time or degree of rotation (Figure 5.7(a) and (b) ). However, this can take some considerable time. For example, it can take several hours of computing time to find all the conform- ations of a simple molecule containg six bonds if energy calculations are made at a rate of 10 determinations per second.

### Quantum mechanics

Unlike molecular mechanics, the quantum mechanical approach to molecular modelling does not require the use of parameters similar to those used in molecular mechanics. It is based on the realization that electrons and all mater- ial particles exhibit wavelike properties. This allows the well defined, parameter free, mathematics of wave motions to be applied to electrons, atomic and molecular structure. The basis of these calculations is the Schrodinger wave equation, which in its simplest form may be stated as:

*H*C ¼ *E*C (5:5)

where C is a mathematical function known as the state function or time- dependent wave function, which defines the state (nature and properties) of a system. In molecular modelling terms *E*C represents the total potential and kinetic energy of all the particles (nuclei and electrons) in the structure and *H* is the Hamiltonium operator acting on the wave function C. Operators are math- ematical methods of converting one function into another function in order to find a solution or solutions of the original function. For example, differentiation is an operator that transforms an equation representing a function into its first derivative.

Schrodinger equations for atoms and molecules use the the sum of the potential and kinetic energies of the electrons and nuclei in a structure as the basis of a description of the three dimensional arangements of electrons about the nucleus. Equations are normally obtained using the Born–Oppenheimer approximation, which considers the nucleus to be stationary with respect to the electrons. This approximation means that one need not consider the kinetic energy of the nuclei in a molecule, which considerably simplifies the calculations. Furthermore, the form of the Schrodinger equations shown in Equation (5.5) is deceptive in that it is not a single equation but represents a set of differential wave equations (C*n*), each corresponding to an allowed energy level (*En*) in the structure. The fact that a structure will only possess energy levels with certain specific values is a direct consequence of spectroscopic observations.

The precise mathematical form of *E* C for the Schrodinger equation will depend on the complexity of the structure being modelled. Its operator *H* will contain individual terms for all the possible electron–electron, electron–nucleus and nucleus–nucleus interactions between the electrons and nuclei in the struc- ture needed to determine the energies of the components of that structure. Consider, for example, the structure of the hydrogen molecule with its four particles, namely two electrons at positions *r*1 and *r*2 and two nuclei at positions *R*1 and *R*2. The Schrodinger Equation (5.5) may be rewritten for this molecule as:

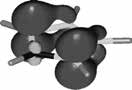
*H*C ¼ (*K* þ *U* )C ¼ *E*C (5:6)

where *K* is the kinetic and *U* is the potential energy of the two electrons and nuclei forming the structure of the hydrogen molecule. The Hamiltonian operator for this molecule will contain operator terms for all the interactions between these particles and so may be written as:

electrons, and the remaining terms represent all the possible interactions

between the relevant electrons and nuclei. The more electrons and nuclei there are in the structure the more complex *H* becomes and as a direct result the greater the computing time required to obtain solutions of the equation. Conse- quently, in practice it is not economic to obtain solutions for structures consist- ing of more than about 50 atoms.

It is not possible to obtain a direct solution of a Schrodinger equation for a structure containing more than two particles. Solutions are normally obtained by simplifying *H* by using the Hartree–Fock approximation. This approxima- tion uses the concept of an effective field *V* to represent the interactions of an electron with all the other electrons in the structure. For example, the Hartree– Fock approximation converts the Hamiltonian operator (5.7) for each electron in the hydrogen molecule to the simpler form: where *r* is the position of the electron. The use of the Hartree–Fock approxima- tion reduces computer time and reduces the cost without losing too much in the way of accuracy. Computer time may be further reduced by the use of semi- empirical methods. These methods use experimentally determined data to sim- plify many of the atomic orbitals, which in turn simplifies the Schrodinger equation for the structure. Solving the Schrodinger equation uses a mathemat- ical method, which is initially based on guessing a solution for each electrons molecular orbital. The computer tests the accuracy of this trial solution and based on its findings modifies the trial solution to produce a new solution. The accuracy of this new solution is tested and a further solution is proposed by the computer. This process is repeated until the testing the solution gives answers within acceptable limits. In molecular modelling the solutions obtained by the use of these methods describe the molecular orbitals of each electron in the molecule. The solutions are normally in the form of sets of equations, which may be interpretated in terms of the probability of finding an electron at specific points in the structure. Graphics programs may be used to convert these prob- abilities into either presentations like those shown in Figures 5.1 and 5.2 or into electron distribution pictures (Figure 5.8). However, because of the computer time involved, it is not feasible to deal with structures with more than several hundred atoms, which makes the quantum mechanical approach less suitable for large molecules such as the proteins that are of interest to medicinal chemists.

H

H H H

N

H

H H H N

H H

Pyrrole Pyrrole, orientation

in the model

The stick picture of pyrrole on which is superimposed the probability of finding electrons at different points in the molecule obtained using quantum mechanics

Quantum mechanics is useful for calculating the values of ionization poten- tials, electron affinities, heats of formation and dipole moments and other physical properties of atoms and molecules. It can also be used to calculate the relative probabilities of finding electrons (the electron density) in a structure. This makes it possible to determine the most likely points at which a structure will react with electrophiles and nucleophiles. A knowledge of the shape and electron density of a molecule may also be used to assess the nature of the binding of a possible drug to a target site (see section 5.5).

### Docking

The three dimensional structures produced on a computer screen may be manipulated on the screen to show different views of the structures. With more complex molecular mechanics programs it is possible to superimpose one structure on top of another. In other words, it is possible to superimpose the three dimensional structure of a potential drug on its possible target site. This process, which is often automated, is known as docking (Figure 5.9). It enables the medicinal chemist to evaluate the fit of potential drugs (ligands) to their target site. If the structure of a ligand is complementary to that of its target site the ligand is more likely to be biologically active. Furthermore, the use of a colour code to indicate the nature of the atoms and functional groups present in the three dimensional structures also enables the medicinal chemist to investi- gate the binding of the ligand to the target site.