17 Computers in medicinal chemistry

Computers are an essential tool in modern medicinal chemistry and are important in both drug discovery and drug development. Rapid advances in computer hardware and software have meant that many of the operations which were once the exclusive province of the expert can now be carried out on ordinary laboratory computers with little specialist expertise in the molecular or quantum mechanics involved. In this chapter, we shall look at examples of how computers are used in medicinal chemistry. However, it has to be appreciated that it is not possible to do full justice to the subject in a single chapter. For example, a full coverage would include details of the mathematics and equations used in different algorithms and that is not possible here. Readers with an interest in gaining a more detailed appreciation of how software programs work at the mathematical level are encouraged to read more specialised textbooks and journal articles (see further reading).

17.1 Molecular and quantum mechanics

The various operations which are carried out in molecular modelling involve the use of programs or **algorithms** which calculate the structure and property data for the molecule in question. For example, it is possible to calculate the energy of a particular arrangement of atoms (conformation), modify the structure to create an energy minimum and calculate properties such as charge, dipole moment and heat of formation. The computational methods that are used to calculate structure and property data can be split into two categories—molecular mechanics and quantum mechanics.

17.1.1 Molecular mechanics

In molecular mechanics, equations are used which follow the laws of classical physics and apply them to nuclei without consideration of the electrons. In essence, the molecule is treated as a series of spheres (the atoms) connected by springs (the bonds). Equations derived from classical mechanics are used to calculate the different interactions and energies (**force fields**) resulting from bond stretching, angle bending, non-bonded interactions and torsional energies. (Torsional energies are associated with atoms that are separated from each other by three bonds. The relative orientation of these atoms is defined by the dihedral or torsion angle—see for example Fig. 17.18.)

These calculations require data or parameters which are stored in tables within the program and which describe interactions between different sets of atoms. The energies calculated by molecular mechanics have no meaning as absolute quantities but are useful when comparing different conformations of the same molecule. Molecular mechanics is fast and is less intensive on computer time than quantum mechanics. However, it cannot calculate electronic properties because electrons are not included in the calculations.

17.1.2 Quantum mechanics

Quantum mechanics uses quantum physics to calculate the properties of a molecule by considering the interactions between the electrons and nuclei of the molecule. Unlike molecular mechanics, atoms are not treated as solid spheres. In order to make the calculations feasible, various approximations have to be made.

- Nuclei are regarded as motionless. This is reasonable since the motion of the electrons is much faster in comparison. As electrons are considered to be moving around fixed nuclei, it is possible to describe electronic energy separately from nuclear energy.
- It is assumed that the electrons move independently of each other, so the influence of other electrons and nuclei is taken as an average.

Quantum mechanical methods can be subdivided into two broad categories—*ab initio* and semi-empirical. The former is more rigorous and does not require any stored parameters or data. It is expensive on computer time, however, and is restricted to small molecules. Semi-empirical methods compute for valence electrons only. They are quicker though less accurate, and can be carried out on larger molecules. There are various forms of semi-empirical software (i.e. programs such as MINDO/3, MNDO, MNDO-d, AM1 and PM3). These methods are quicker, because they use further approximations and make use of stored parameters.

17.1.3 Choice of method

The method of calculation chosen depends on what calculation needs to be done, as well as the size of the molecule. As far as size of molecule is concerned, *ab initio* calculations are limited to molecules containing tens of atoms, semi-empirical calculations on molecules containing hundreds of atoms and molecular mechanics on molecules containing thousands of atoms.

Molecular mechanics is useful for the following operations or calculations:

- energy minimization
- identifying stable conformations
- energy calculations for specific conformations
- generating different conformations
- studying molecular motion.

Quantum mechanical methods are suitable for calculating the following:

- molecular orbital energies and coefficients
- heat of formation for specific conformations
- partial atomic charges calculated from molecular orbital coefficients
- electrostatic potentials
- dipole moments
- transition state geometries and energies
- bond dissociation energies.

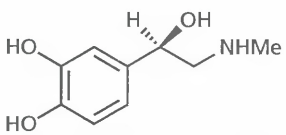
17.2 Drawing chemical structures

Chemical drawing packages do not require the calculations described in section 17.1, but they are often integrated into molecular modelling programs. Various software packages such as ChemDraw, ChemWindow and Isis/Draw are available that can be used to construct diagrams quickly and to a professional standard. For example, the diagrams in this book have all been prepared using the ChemDraw package.

Some drawing packages are linked to other items of software which allow quick calculations of various molecular properties. For example, the following properties for **adrenaline** were obtained using ChemDraw Ultra; the structure's correct IUPAC chemical name, molecular formula, molecular weight, exact mass, and theoretical elemental analysis. It was also possible to get calculated predictions of the compound's ^{1}H and ^{13}C NMR chemical shifts, melting point, freezing point, log *P* value, molar refractivity, and heat of formation (Fig. 17.1).

17.3 3D structures

Molecular modelling software allows the chemist to construct a three-dimensional (3D) molecular structure on the computer. There are several software packages available such as Chem 3D, Alchemy, Sybyl, Hyperchem, Discovery Studio Pro and CAChe. The 3D model can be made by constructing the molecule atom by atom, and bond by bond. It is also possible to automatically convert a 2D drawing into a 3D structure, and most molecular modelling packages have this facility. For example, the 2D structure of adrenaline in Fig. 17.2 was drawn in ChemDraw, then copied and

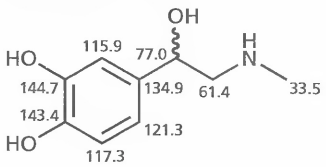


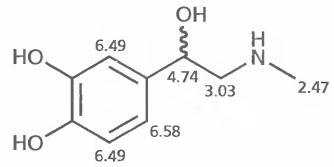
1-(3,4-Dihydroxyphenyl)-2-methylaminoethanol

Calculated Properties
$C_9H_{13}NO_3$
Exact Mass: 183.09
Mol. Wt.: 183.20
C, 59.00; H, 7.15; N, 7.65; O, 26.20

Predicted properties
LogP = –0.61–0.63
Molar refractivity 48.66–49.08 [cm.cm.cm/mol]
b.pt. 618.55 K; Freezing point 539.03 K
Heat of formation –451.22 kJ/mol

Predicted ^{13}C nmr





Predicted ^{1}H nmr

FIGURE 17.1 Drawing chemical structures.

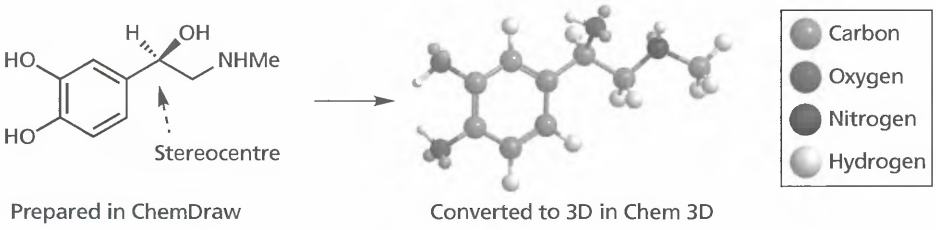


FIGURE 17.2 Conversion of a two-dimensional drawing to a three-dimensional model.

pasted into Chem 3D, resulting in the automatic construction of the 3D model shown. The 3D structures of a large number of small molecules can also be accessed from the Cambridge Structural Database (CSD) and downloaded so that they can be studied. This database contains over 200 000 molecules, which have been crystallized and their structure determined by X-ray crystallography.

17.4 Energy minimization

Whichever software program is used to create a 3D structure, a process called energy minimization should be carried out once the structure is built. This is because the construction process may have resulted in unfavourable bond lengths, bond angles or torsion angles. Unfavourable non-bonded interactions may also be present (i.e. atoms from different parts of the molecule occupying the same region of space). The energy minimization process is usually carried out by a molecular mechanics program which calculates the energy of the starting molecule, then varies the bond lengths, bond angles and torsion angles to create a new structure. The energy of the new structure is calculated to see whether it is energetically more stable or not. If the starting structure is inherently unstable, a slight alteration in bond angle or bond length will have a large effect on the overall energy of the molecule resulting in a large energy difference (ΔE; Fig. 17.3). The program will recognize this and carry out more changes, recognizing those which lead to stabilization and those which do not. Eventually, a structure will be found where structural variations result in only slight changes in energy—an energy minimum. The program will interpret this as the most stable structure and will stop at that stage (Box 17.1).

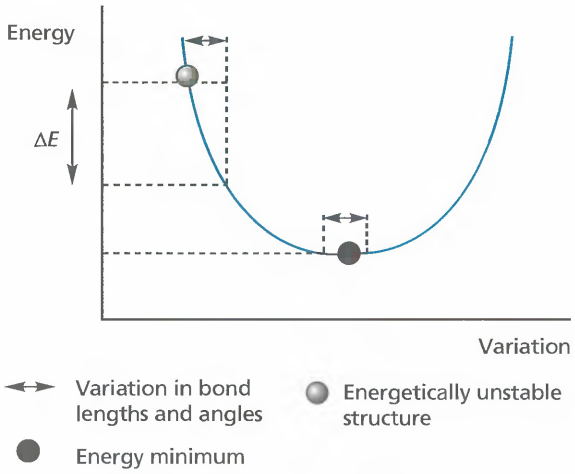


FIGURE 17.3 Energy minimization.

BOX 17.1 Energy minimizing apomorphine

A 2D structure of **apomorphine** was converted to a 3D structure using Chem 3D. However, the catechol ring was found to be non-planar with different lengths of C–C bond. Energy minimization corrected the deformed aromatic ring, resulting in the desired planarity and the correct length of bonds.

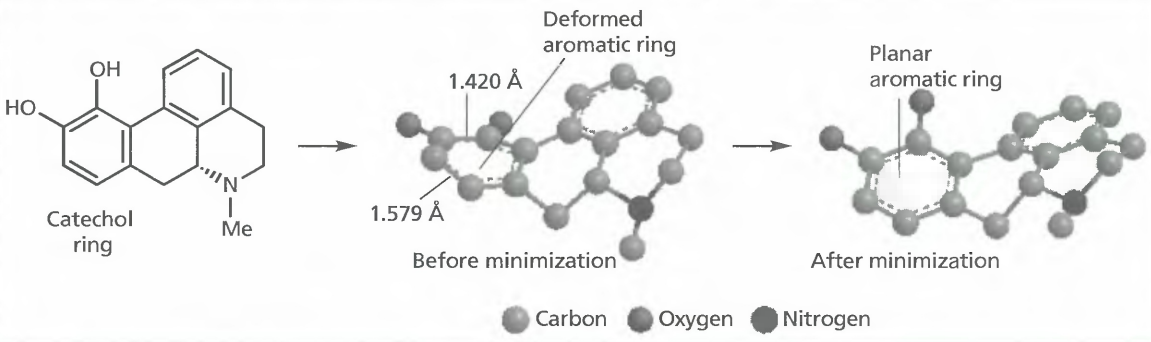


FIGURE 1 Energy minimization carried out on apomorphine.

17.5 Viewing 3D molecules

Once a structure has been energy minimized, it can be rotated in various axes to study its shape from different angles. It is also possible to display the structure in different formats (i.e. cylindrical bonds, wire frame, ball and stick, space filling; Fig. 17.4).

There is another format known as the ribbon format which is suitable for portraying regions of protein secondary structure such as α-helices. This often simplifies the highly complex-looking structure of a protein, allowing easier visualization of its secondary and tertiary structure. The ball and stick model of an α-helical decapeptide consisting of 10 alanine units is shown in Fig. 17.5, along with the same molecule displayed as a ribbon.

17.6 Molecular dimensions

Once a 3D model of a structure has been constructed, it is a straightforward procedure to measure all of its bond lengths, bond angles and torsion (or dihedral) angles. These values can be read from tables or by highlighting the relevant atoms and bonds on the structure itself. The various bond lengths, bond angles and torsion angles measured for **adrenaline** are illustrated in Fig. 17.6. It is also a straightforward process to measure the separation between any two atoms in a molecule.

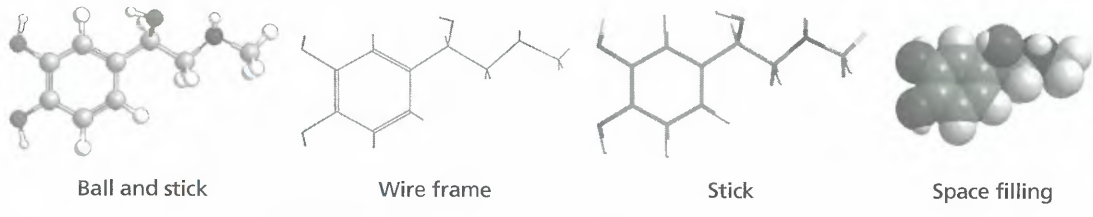


FIGURE 17.4 Different methods of visualizing molecules.

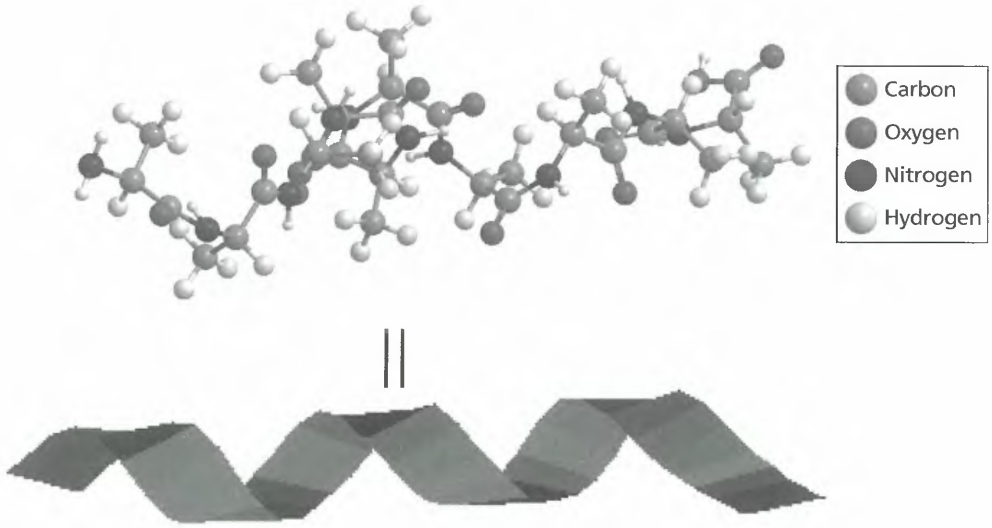


FIGURE 17.5 Ribbon representation of a helical decapeptide (Chem 3D).

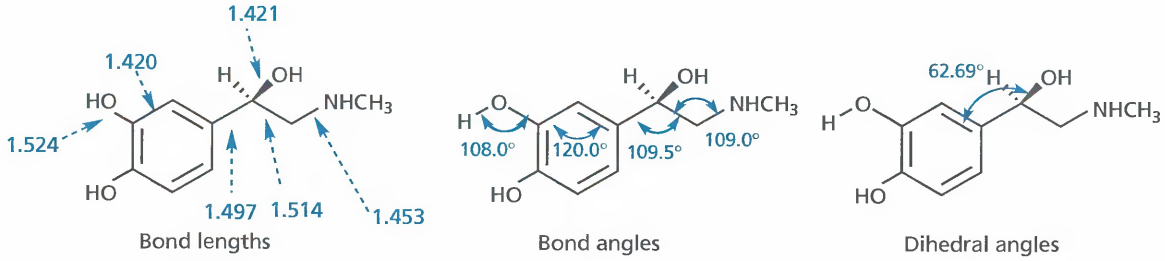


FIGURE 17.6 Molecular dimensions for adrenaline (Chem 3D).

17.7 Molecular properties

Various properties of the 3D structure can be calculated once it has been built and minimized. For example, the steric energy is automatically measured as part of the minimization process and takes into account the various strain energies within the molecule, such as bond stretching or bond compression, deformed bond angles, deformed torsion angles, non bonded interactions arising from atoms too close to each other in space, and unfavourable dipole dipole interactions. The steric energy is useful when comparing different conformations of the same structure, but the steric energies of different molecules should not be compared.

Other properties for the structure can be calculated, such as the predicted heat of formation, dipole moment, charge density, electrostatic potential, electron spin density, hyperfine coupling constants, partial charges, polarizability, and infrared vibrational frequencies. Some of these are described below.

17.7.1 Partial charges

It is important to realize that the valence electrons in molecules are not fixed to any one particular atom and can move around the molecule as a whole. As the electrons are likely to spend more of their time nearer electronegative atoms than electropositive atoms, this distribution is not uniform and results in some parts of the molecule being slightly positive and other parts being slightly negative. For example, the partial charges for **histamine** are shown in Fig. 17.7.

The calculation of partial charges has important consequences in the way we view ions. Conventionally, we consider charges to be fixed on a particular atom (unless delocalization is possible). For example, the histamine ion is normally drawn showing the positive charge on the terminal nitrogen atom (Fig. 17.8). In fact, calculation of partial charges shows that some of the positive charge is localized on the hydrogens attached to the terminal nitrogen. This has important consequences in the way we think of ionic interactions between a drug and its binding site. It implies that charged areas in the binding site and the drug are more diffuse than one might think. This in turn suggests that we have wider scope in designing novel drugs. For example, in the classical viewpoint of charge distribution, a certain molecule might be considered to have its charged centre too far away from the corresponding 'centre' in the binding site. If these charged areas are actually more diffuse, then this is not necessarily true (Fig. 17.9).

It is worth pointing out, however, that such calculations are carried out on structures in isolation from their environment. Histamine is in an aqueous environment in the body, and would be surrounded by water molecules

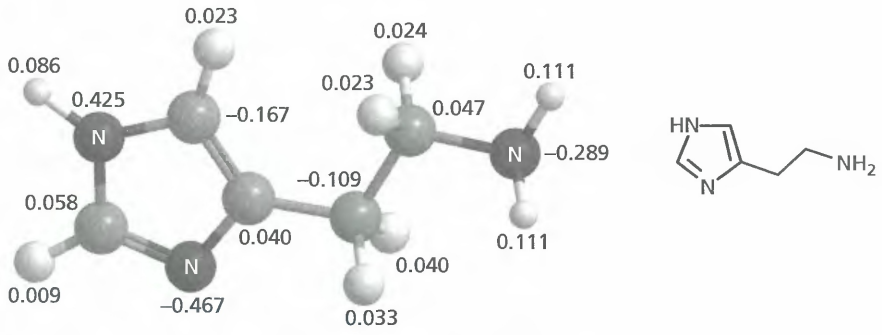


FIGURE 17.7 Partial charges for histamine.

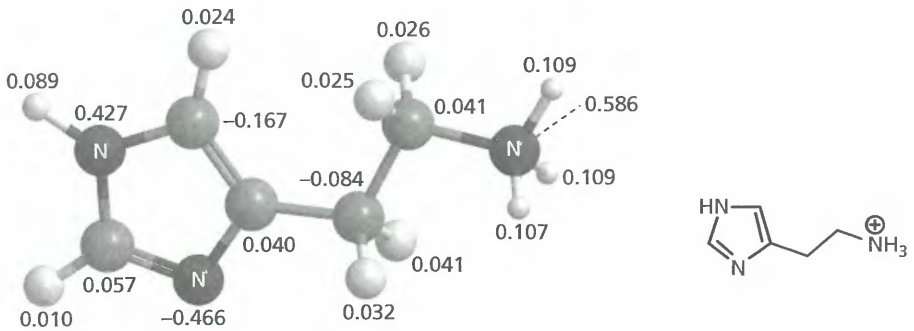


FIGURE 17.8 Charge distribution on the histamine ion.

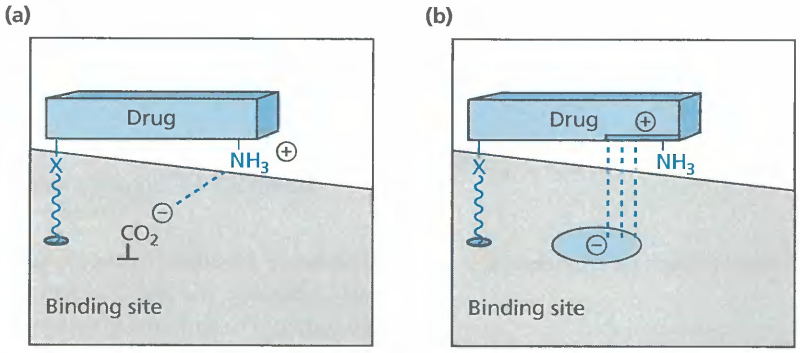


FIGURE 17.9 Ionic interactions. **(a)** Classic view of an ionic interaction and **(b)** diffuse ionic interaction.

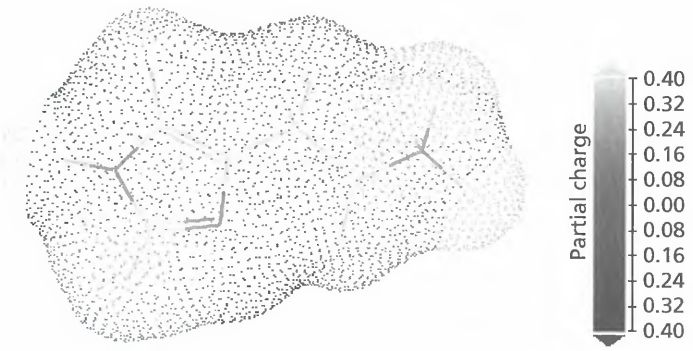


FIGURE 17.10 Dot cloud representation of partial charge.

which would solvate the charge and consequently have an effect on charge distribution. Furthermore, water has a high dielectric constant, which means that electrostatic interactions are more effectively masked than in a hydrophobic environment.

Partial charges can also be represented by dot clouds. The size of each cloud represents the amount of charge and the clouds can be coloured to show what sort of charge it is. The dot clouds shown in Fig. 17.10 represent the relative sizes of partial charges on the histamine ion.

17.7.2 Molecular electrostatic potentials

Another way to consider charge distribution is to view the molecule as a whole rather than as individual atoms and bonds. This allows one to identify *areas* of the molecule which are electron rich or electron poor. This is particularly important in the **3D QSAR** technique of **CoMFA** described in section 18.10. It can also be useful in identifying how compounds with different structures might line up to interact with corresponding electron rich and electron poor areas in a binding site.

Molecular electrostatic potentials (MEPs) can be calculated using quantum mechanics by considering the molecular orbitals. The MEP for histamine shown in Fig. 17.11 was calculated using the semi-empirical method AM1. Another method of calculating MEPs is described in section 17.7.5.

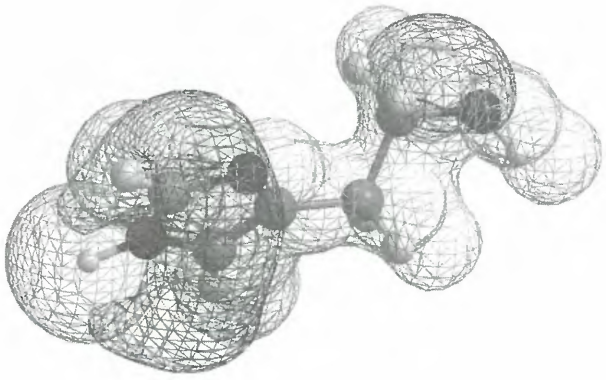


FIGURE 17.11 Molecular electrostatic potential for histamine.

An example of how electrostatic potentials have been used in drug design can be seen in the design of the **cromakalim** analogue (II; Fig. 17.12), where the cyanoaromatic ring was replaced by a pyridine ring. This was part of a study looking into analogues of cromakalim which

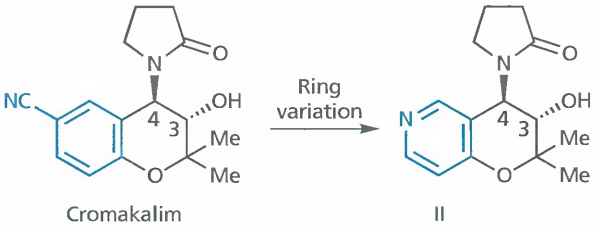


FIGURE 17.12 Ring variation on cromakalim.

would have similar antihypertensive properties, but which might have different pharmacokinetics. In order to retain activity, it was important that any replacement heteroaromatic ring was as similar in character to the original aromatic ring as possible. Consequently, the MEPs of various bicyclic systems were calculated and compared with the parent bicyclic system (III; Fig. 17.13). In order to simplify the analysis, the study was carried out in 2D within the plane of the bicyclic systems, and maps were created showing areas of negative potential (Fig. 17.14). The contours represent the various levels of the MEP, and can be taken

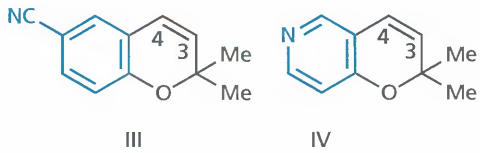


FIGURE 17.13 Bicyclic models in cromakalim study

to indicate possible hydrogen bonding regions around each molecule. The analysis demonstrated that the bicyclic system (IV) had similar electrostatic properties to (III), resulting in the choice of structure (II) as an analogue.

17.7.3 Molecular orbitals

The molecular orbitals of a compound can be calculated using quantum mechanics. For example, **ethene** can be shown to have 12 molecular orbitals. The **highest occupied molecular orbital (HOMO)** and **lowest unoccupied molecular orbital (LUMO)** are shown in Fig. 17.15 (see also Box 17.2).

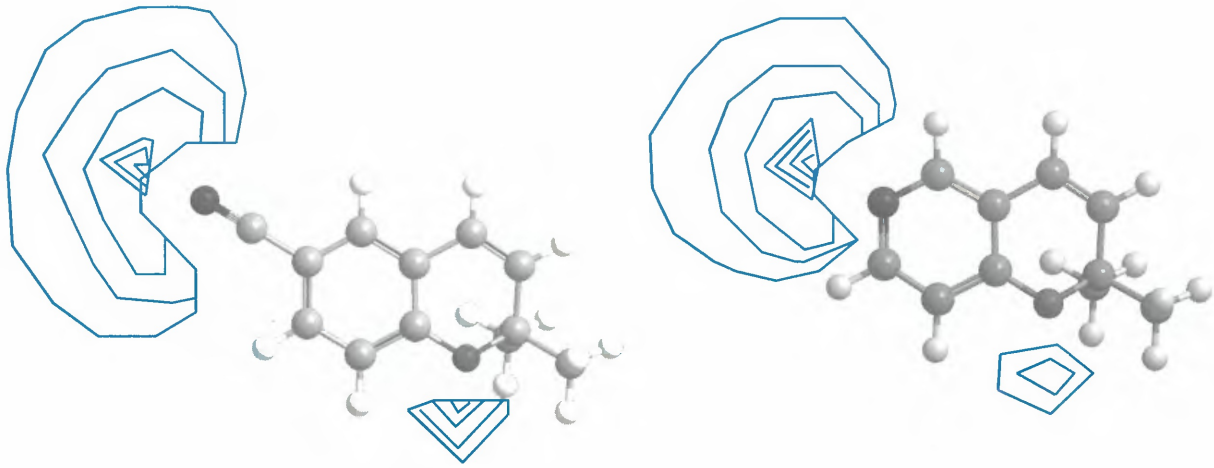


FIGURE 17.14 MEPs of bicyclic models (III) and (IV).

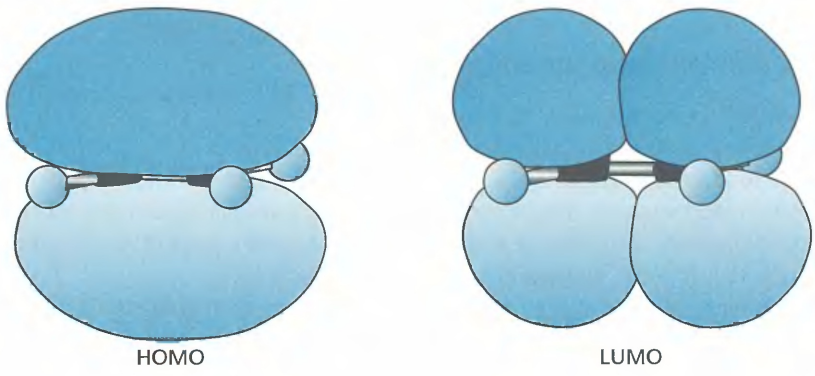


FIGURE 17.15 HOMO and LUMO molecular orbitals for ethene.

BOX 17.2 Study of HOMO and LUMO orbitals

A study of a molecule's HOMO and LUMO orbitals is useful because frontier molecular orbital theory states that these orbitals are the most important in terms of a molecule's reactivity. HOMO and LUMO orbitals can also help to explain drug–receptor interactions. For example, **ketanserin** (Fig. 1) is an antagonist at serotonin receptors, but has a greater binding affinity than would be expected from the more obvious intermolecular interactions.

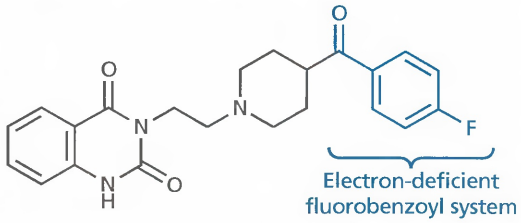


FIGURE 1 Ketanserin.

In order to explain this greater binding affinity, it was proposed that a charge transfer interaction was taking place between the electron-deficient fluorobenzoyl ring system of ketanserin and an electron-rich tryptophan residue, which was known to be nearby in the binding site. To check this, HOMO and LUMO energies were calculated for a model complex between the indole system of tryptophan and the fluorobenzoyl system of ketanserin (Fig. 2). This showed that the HOMO for the indole/fluorobenzoyl complex resided on the indole structure whereas the LUMO was on the fluorobenzoyl moiety, indicating that charge transfer was possible. With other antagonists, there was not this same clear-cut separation between the HOMO and LUMO orbitals, with the indole system being involved in both orbitals.

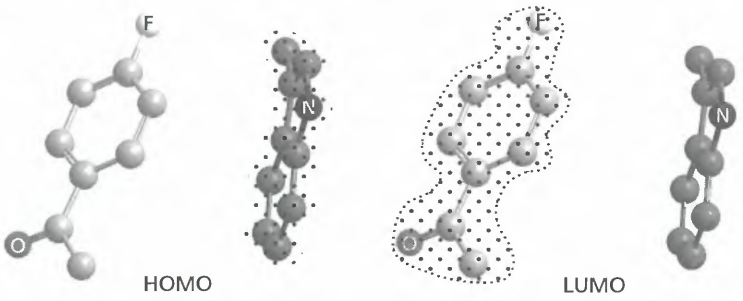


FIGURE 2 HOMO and LUMO molecular orbitals (dot surfaces) for the indole/fluorobenzoyl complex.

17.7.4 Spectroscopic transitions

It is possible to calculate the infrared or ultraviolet transitions for a molecule. Although a theoretical infrared spectrum can be generated, it is highly unlikely that it will accurately match the actual infrared spectrum. Nevertheless, the position and identification of specific absorptions can be identified and can be useful in the design of drugs. For example, it is found that the activity of penicillins is related to the position of the β-lactam carbonyl stretching vibration in the infrared spectrum. Calculating the theoretical wavenumber for a range of β-lactam structures can be useful in identifying which ones are likely to have useful activity before synthesizing them.

17.7.5 Use of grids in measuring molecular properties

Grids have become extensively used in measuring molecular properties, and are important in a variety of software programs used both in docking (section 17.12) and 3D-QSAR (section 18.10).

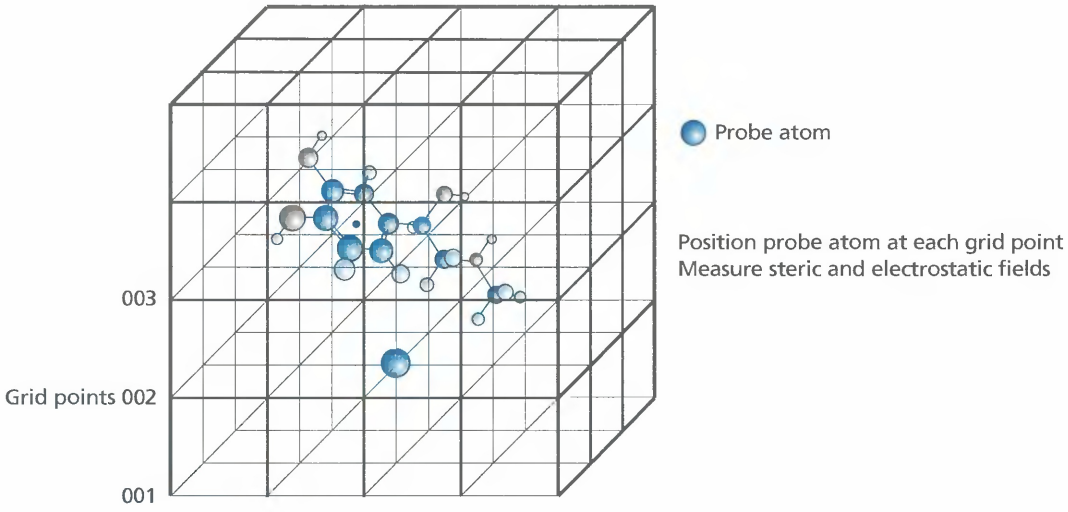


FIGURE 17.16 Measuring fields around a molecule by placing a probe atom at grid points.

There are various properties of a molecule which can be measured as **fields**. A field is defined as the influence that a property has on the space surrounding the molecule. As an analogy, consider a magnet. This creates a magnetic field around it which gets stronger the closer one gets to the magnet. In the same way, it is possible to measure a molecular property by the influence it has on surrounding space. The most commonly measured molecular fields are steric and electrostatic. These can be measured by placing a molecule into a preconstructed 3D lattice or grid (Fig. 17.16). The intersections of this lattice are called lattice (or grid) points and these define the 3D space around the molecule.

Once a molecule has been placed into the lattice, the steric and electrostatic fields around it can be measured. This is done by placing a probe atom such as a proton or an sp^3 hybridized carbocation at each of the grid points in turn, and using software to calculate the steric and electrostatic interactions between the probe and the molecule. As far as the steric field is concerned, this will increase as the probe atom gets closer to the molecule. As far as the electrostatic field is concerned, there will be an attraction between the positively charged probe and electron rich regions of the molecule, and a repulsion between the probe and electron deficient regions of the molecule.

The steric and electrostatic fields at each grid point are tabulated, and a particular value for the steric energy is chosen which will define the shape of the molecule. The grid points having that value are then connected by contour lines to define the steric field. A similar process is carried out to measure the electrostatic interactions between the positively charged probe atom and the test molecule. Electron rich and electron deficient regions for the molecule are then defined by suitable contour lines. It is also possible to use the grid method to measure a hydrophobic field by using a water molecule as a probe.

Grids can also be constructed within the binding sites of enzymes or receptors, and are important in many docking software programs (section 17.12). Various atoms or molecular fragments are used as probes in order to measure interactions with the amino acids that make up the molecular surface of the binding site. The interactions of interest are typical binding interactions such as ionic, van der Waals and hydrogen bonding. The atoms or fragments used as probes are the typical atoms or fragments that might be found in a drug molecule. For example typical atom probes are C, H, N and O. Typical fragment probes are C=O, CO_2^-, N–H etc. In this way, it is possible to measure whether various binding interactions are possible at the different grid points, as well as their strength. The measurements can then be stored in tables for each atom or molecular fragment. This obviously involves a lot of calculations, but it only has to be done once in order to define the binding characteristics of the binding site. Once this has been done the binding strengths of different docked molecules can be quickly calculated by identifying which atoms or groups coincide with particular grid points. The relevant entries in the tables are 'looked up' and summed to give the required total. In this way it is possible to carry out docking studies on hundreds of different molecules within a reasonable time span. The use of grids to measure the binding characteristics of an enzyme active site was important in the development of the anti-flu drug **zanamivir** (section 20.8.3.2).

KEY POINTS

- Several chemical drawing packages include software allowing the calculation of various physical properties.
- Molecular modelling software makes use of programs based on molecular mechanics and quantum mechanics.
- Molecular mechanics programs use equations based on classical physics to calculate force fields. Atoms are treated as spheres, bonds as springs. Electrons are ignored. This method is suitable for energy minimization and conformational analysis.
- Quantum mechanical methods are *ab initio* or semi-empirical. The former is more rigorous but is restricted to small molecules. These methods are suitable for measuring molecular properties such as molecular orbital energies and coefficients.
- Energy minimization has to be carried out on any molecule constructed with molecular modelling software. The process involves alteration of bond lengths, bond angles, torsion angles, and non-bonded interactions until a stable conformation is obtained.
- Molecular modelling software allows the dimensions of a molecule to be accurately measured as well as its partial charges, molecular electrostatic potentials, and molecular orbitals.
- Grids and probe atoms are used to measure steric, electrostatic and hydrophobic fields around molecules.
- Grids can be placed in binding sites in order to identify the nature and strength of potential binding interactions at different locations within the binding site. These can be tabulated and used to measure binding energies of ligands.

17.8 Conformational analysis

17.8.1 Local and global energy minima

In section 17.4, we saw how energy minimization is carried out on a 3D structure to produce a stable conformation. The structure obtained, however, is not necessarily the most stable conformation. This is because energy minimization stops as soon as it reaches the first stable conformation it finds, and that will be the one closest to the starting structure. This is illustrated in Fig. 17.17 where the most stable conformation is separated from a less stable conformation by an energy saddle. If the 3D structure initially created is on the energy curve at the position shown, energy minimization will stop when it reaches the first stable conformation it encounters—**a local energy minimum**. At this point, variations in structure result in low-energy changes and so the minimization will stop. In order to cross the saddle to the more stable conformation, structural variations would have to be carried out which increase the strain energy of the structure and these will be rejected by the program. The minimization program has no way of knowing that there is a more stable conformation (**a global energy minimum**) beyond the energy saddle. Therefore, in order to identify the most stable conformation, it is necessary to generate different conformations of the molecule and to compare their steric energies. We shall now look at some of the ways in which this can be done.

17.8.2 Molecular dynamics

Molecular dynamics is a molecular mechanics program designed to mimic the movement of atoms within a molecule. The software program works by treating the atoms in the structure as moving spheres. After one femtosecond (1×10^{-15} s) of movement, the position and velocity of each atom in the structure is determined. The forces acting on each atom are then calculated by considering bond lengths, bond angles, torsional terms and non bonded interactions with surrounding atoms. The potential energy of each atom is calculated and Newton's laws of motion are then used to determine the acceleration and direction of movement of each atom (the kinetic energy). This allows the program to predict the velocity and position of each atom a femtosecond later. The procedure is then repeated for every femtosecond of the process. The femtosecond duration is important, as it is an order of magnitude less than the rate of a bond stretching vibration, and so an atom is only allowed to move a fraction of a bond length between each calculation. If this was not the case, and atoms were allowed to move greater distances, one might get the situation where two atoms occupy the same area of space. The calculated forces

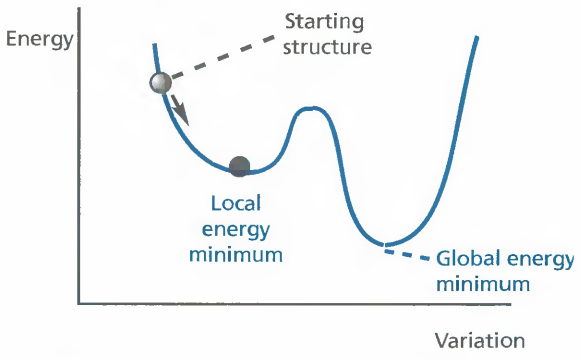


FIGURE 17.17 Local and global energy minima.

and potential energies would then be huge leading to atoms moving with excessive velocities and accelerations and the system would fail.

Molecular dynamics can be used to generate a variety of different conformations by 'heating' the molecule to 900 K. Of course, this does not mean that the inside of your computer is about to melt! It means that the program allows the structure to undergo bond stretching and bond rotation as if it *was* being heated. As a result, energy barriers between different conformations are overcome, allowing the crossing of energy saddles. In the process, the molecule is 'heated' at a high temperature (900 K) for a certain period (e.g. 5 ps), then 'cooled' to 300 K for another period (e.g. 10 ps) to give a final structure.

The process can be repeated automatically as many times as desired, to give as many different structures as is practical. Each of these structures can then be recovered, energy minimized, and its steric energy measured. By carrying out this procedure, it is usually possible to identify some distinct conformations that might be more stable than the initial conformation.

For example, the two-dimensional drawing of **butane** shown in Fig. 17.18 was imported into Chem-3D and energy minimized. Because of the way the molecule was represented, energy minimization stopped at the first local energy minimum it found—the *gauche* conformation having a steric energy of 3.053 kcal/mole. The molecular dynamics program was run to generate other conformations and successfully produced the fully staggered *anti* conformation which, after minimization, had a steric energy of 2.180 kcal/mole (i.e. it was more stable by about 0.9 kcal/mole).

In fact, this particular problem could be solved more efficiently by the stepwise rotation of bonds described below (section 17.8.3). Molecular dynamics is more useful for creating different conformations of molecules which are not conducive to stepwise bond rotation (e.g. cyclic systems—Box 17.3), or which would take too long to analyse by that process (large flexible molecules).

Finally, it has to be remembered that biomolecules in the real world are surrounded by water and that this can affect the relative stability of different conformations. Therefore, it is advisable to include water molecules in the modelling system before carrying out molecular dynamics experiments.

17.8.3 Stepwise bond rotation

Although molecular dynamics can be used to generate different conformations, there is no guarantee that it will identify all the conformations which are possible for a structure or find the global minimum. A more systematic process is to generate different conformations by automatically rotating every single bond by a set number of degrees. For example, 12 different conformations of butane were generated by automatically rotating the central bond in 30° steps. The steric energy of each conformation was calculated and graphed (Fig. 17.19), revealing that the most stable conformation was the fully staggered one, and the least stable conformation was the eclipsed one. In this operation, energy minimization is not carried out on each structure, because the aim is to identify both stable and unstable conformations.

Some modelling software packages can automatically identify all the rotatable single bonds in a structure. Bonds to hydrogen or to methyl groups are excluded in this analysis, as rotations of these bonds do not generate significantly different conformations. Once the rotatable bonds have been identified, the program generates all the possible conformations which can arise from rotating these bonds by a set amount determined by the operator. The number of conformations generated will depend on the number of rotatable bonds present and the set amount of rotation. For example, a structure with three rotatable bonds could be analysed for conformations resulting from 10° increments at each bond to generate 46 656 conformations. With four rotatable bonds, 30° increments would generate 20 736 conformations.

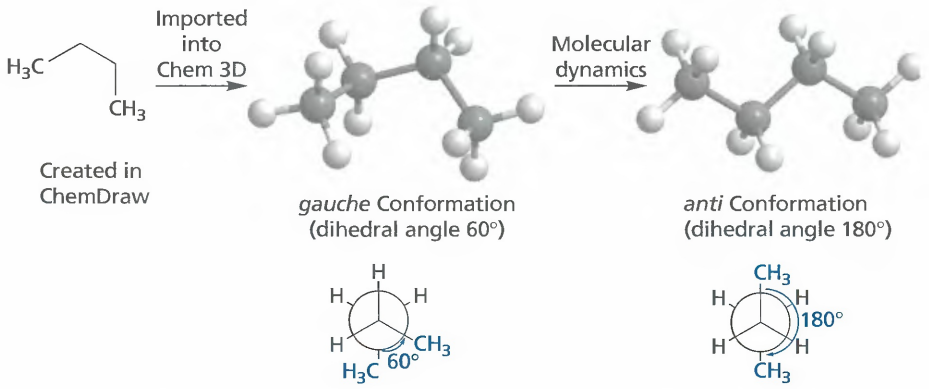
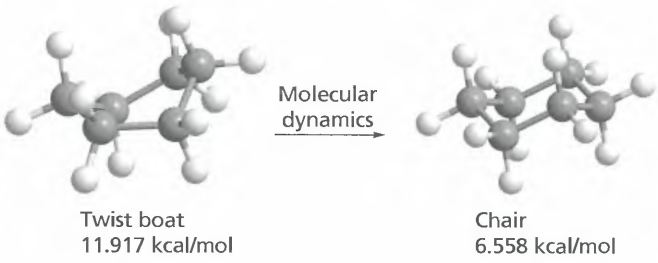


FIGURE 17.18 Use of molecular dynamics to find the most stable conformation.

BOX 17.3 Finding conformations of cyclic structures by molecular dynamics

The twist boat conformation of **cyclohexane** is not the most stable conformation of cyclohexane, but remains as the twist boat when energy minimization is carried out. 'Heating' the molecule by molecular dynamics produces a variety of different conformations including the more stable chair conformation.



Generation of the cyclohexane chair conformation by molecular dynamics in Chem 3D.

In general, about 1000 conformations per second can be processed on a standard bench top computer. However, it is important to be as efficient as possible and care should be taken in deciding how much each bond should be rotated at a time to ensure that a representative but manageable number of conformations is created.

It is possible to make the process more efficient depending on the information desired. For example, if we were only interested in identifying stable conformations, the program can automatically filter out conformations which are eclipsed or near eclipsed. It is also possible to filter out 'nonsense' conformations (i.e. conformations where some atoms occupy the same position in space). Such conformations can arise because bond rotations are being carried out by the program without analysing what is happening elsewhere in the molecule.

Once a series of conformations have been generated, they can be tabulated and sorted into their order of stability. The most stable conformations can then be energy minimized and their structures compared.

17.8.4 Monte Carlo and the Metropolis method

In molecular dynamics, a search for the most stable conformation involves the generation of random

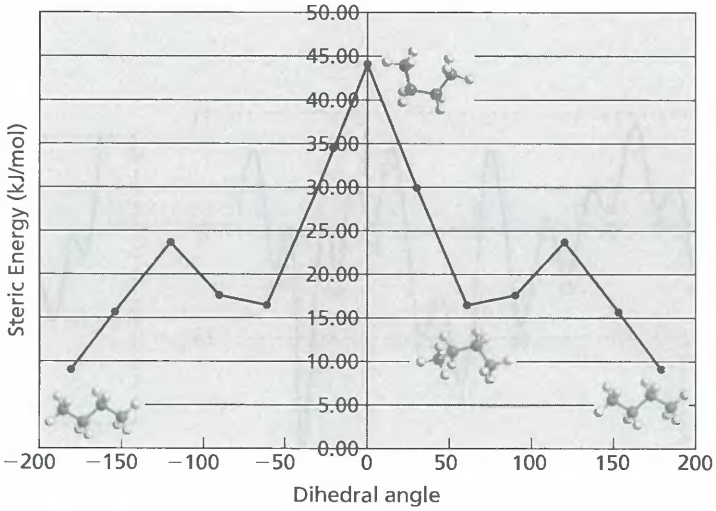


FIGURE 17.19 Graph showing relative stabilities of various butane conformations.

conformations which are all analysed separately. This means that the same amount of processing time is taken up by high-energy and low-energy conformations. The Monte Carlo method of conformational analysis introduces a bias towards stable conformations such that more processing time is spent on these—a process known as importance sampling. Different conformations are generated by carrying out random bond rotations. This is quite different from molecular dynamics where atoms are shifted in space. As each conformation is generated, it is energy minimized to give a stable conformation, and its steric energy is calculated and compared with the previous structure. If the steric energy of the new conformation is lower (more stable), it is accepted and used as the starting structure for the next conformation. If the steric energy is higher, it may be accepted or rejected depending on a probability formula which takes into account both the energy of the new conformation and the 'temperature' of the system. For example, suppose conformation G (Fig. 17.20) is the starting conformation. The new conformation that is generated will be structurally similar to G and could be conformation F or H. Conformation H is more stable and would be immediately accepted and used to generate the next conformation. Conformation F is less stable and so the probability equation would be used to determine whether it is accepted or not. If it is not, another conformation would be generated from conformation G.

The temperature is set by the user and if it is set high enough (for example T1), virtually all the stable conformations that are generated will be accepted and used as the starting point for the next structure. The process is repeated for as long as the user wishes, in order to generate a set of different conformations. The advantage of using a high temperature is that it will allow the algorithm to generate a structurally diverse set of conformations such as conformations B, D, F, G, H, J, K, L and N in figure 17.20. The steric energies can be measured allowing the identification of the most stable conformation present. However, there is no guarantee that the global minimum will be found—in this case conformation M—due to the random nature of the search. This is particularly the case with molecules having several rotatable bonds and a huge number of possible stable conformations.

If this is the case, one might ask why the algorithm could not be run at a lower temperature such that the algorithm only accepts structures that are more stable than the previous one. Unfortunately, this will only mean that the system will focus on a particularly localised area of conformational space. (Conformational space is the term used to describe the various conformations that are possible for a structure.) For example, starting from conformation G, conformations F and H are likely to be generated because they are similar structures, but only H will be accepted. Modifying H may generate conformation I, but there the process will stop. Although structures C and M are more stable, they will not be generated since the search would have to accept higher energy conformations (such as K) in order to create them. In other words, the algorithm will find the most stable conformation closest in nature to the starting structure.

The Metropolis method (also known as **simulated annealing**) is an approach which can be used to increase the chances of finding the global minimum. It involves a number of cycles where the Monte Carlo algorithm is run at different temperatures. In the first cycle, a high temperature is set (T1) and a set of structurally diverse conformations is generated. The most stable conformation is then used as the starting structure for the next run where the temperature is set at a lower value. This process is repeated several times with the probability equation becoming more 'choosy' about which structures are accepted. This slowly 'focuses' the search on a particular

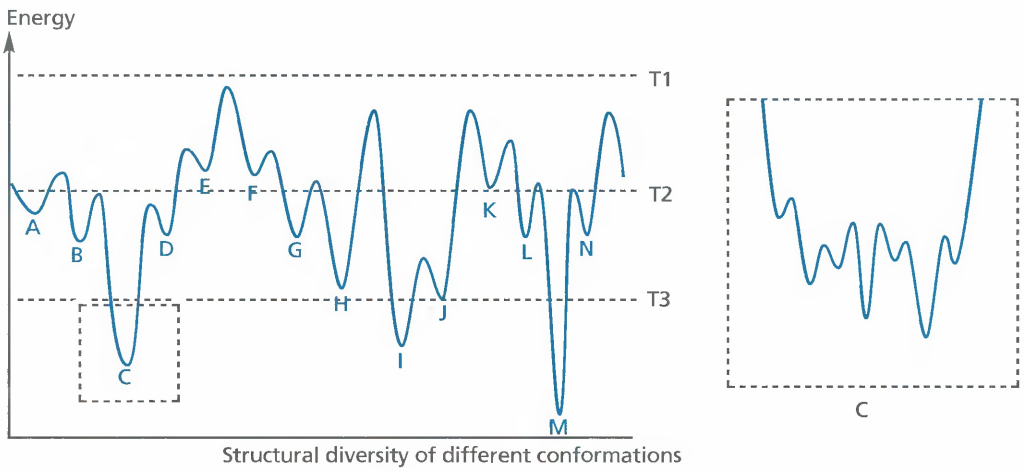


FIGURE 17.20 A two-dimensional representation of conformational space versus structural diversity of conformations.

area of conformational space which can be searched more rigorously. In this way, there is more chance of finding the global minimum, but there is still no guarantee.

For example, an initial run at T1 may generate conformations B, D, F, G, H, J, K, L and N. Conformation J would be taken as the starting conformation for the next run which would be run at a lower temperature (T2). At this temperature, conformations such as E, F and K are less likely to be accepted. The second run might then generate conformations A, C, D, G, H and I but fail to generate conformations L, M and N since they are structurally different and would only be generated if structure K was generated first. This shows how the search is starting to narrow into specific areas of conformational space. Structure C would be the starting point for the next run at a lower temperature (T3) and this would focus the search into even finer detail where very similar conformations are identified and their steric energies compared, leading to the identification of the most stable conformation in that area of conformational space.

Ideally, the more slowly the temperature is lowered and the more runs that are carried out, the more chance that the global minimum will be identified. For example, just using three runs as described above narrows the search into an area which does not include the global minimum M. By decreasing the temperature in smaller increments, there is more chance that the search will focus into the correct area of conformational space, in order to identify the global minimum.

17.8.5 Genetic and evolutionary algorithms

A search for stable conformations can be carried out using genetic and evolutionary algorithms. As the names suggest, these algorithms are programmed to work using the same principles as biological evolution. Consider for example the growth of a bacterial cell. For a cell to divide, DNA has to be copied. However, the copying process is not perfect. Random mutations take place to the genes coding for the bacterial proteins. These mutations may be advantageous or disadvantageous to the individual cell concerned. For example, a mutation may be advantageous if it provides the bacterial cell with immunity against an antibacterial agent. As a result, this cell will survive and the mutation will be passed on to future generations. On the other hand, a mutation might disrupt the function of a vital protein which results in cell death. Consequently, this mutation is not preserved. When we move to the human level, each member of a new generation receives chromosomes from two parents and this provides variety where particular characteristics are received from one parent or the other.

As far as conformational analysis is concerned, genetic and evolutionary algorithms are designed to create

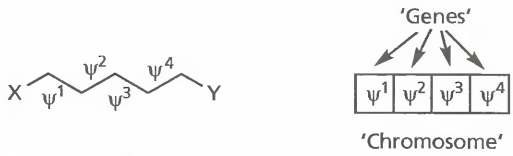


FIGURE 17.21 The representation of torsion angles as genes in chromosomes.

different conformations and to carry out an evolutionary process which will select the most stable conformations. We will now look more closely at how this is done.

First of all, the conformation of a molecule has to be represented in a manner which will allow an evolutionary process of mutation and selection to take place. Quite simply, the torsion angles for the rotatable bonds in the molecule are stored as a sequence of numbers. This sequence corresponds to a 'chromosome', where each 'gene' signifies a torsion angle (Fig. 17.21).

An initial population of 'chromosomes' representing different conformations is created by randomly choosing values for the different torsion angles. The stability of each conformation is then calculated by molecular mechanics. The next stage in genetic algorithms is to create a new population of chromosomes or conformations. First of all, sets of 'parents' are chosen from the initial population. This is a random process, but a statistical bias is built into the selection process such that the most stable conformations are chosen as parents. This means that a particularly stable conformation can be involved in several 'relationships'.

The new population of conformations is now generated. The chromosomes from each parent undergo a 'cross over' or recombination process to generate new chromosomes where each chromosome has torsion angles contributed from each parent. In the example shown below (Fig. 17.22), the cross over involves the first two torsion angles of each chromosome. This generates two 'children' conformations, one of which corresponds to a more stable conformation where all the torsion angles are at 180°.

As well as cross overs, random 'single point' mutations are made on individual chromosomes. This corresponds to a random alteration of a single torsion angle within the chromosome. Both cross over processes and mutations generate a new and diverse population of conformations which can now act as parents for the next generation. The process can be repeated for as long as is practical.

There is a risk that a particularly stable conformation might be formed early on in the process and be lost as a result of further cross overs and mutations. To guard against this, most programs carry forward the chromosomes representing the most stable conformations unchanged—a so called 'elitist' strategy.

In the above example, the torsion angles in the chromosomes were represented by the actual value of the torsion angle. Many programs store these values as a binary

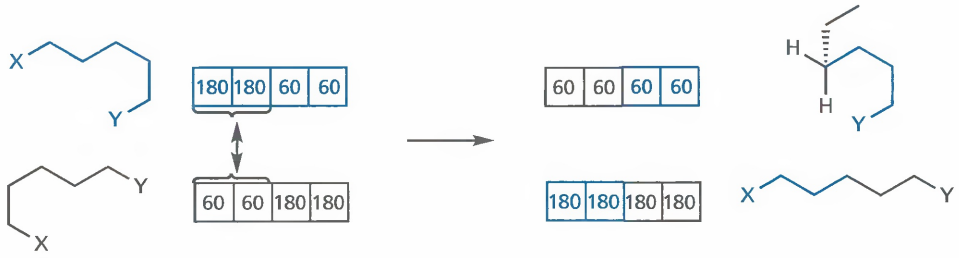


FIGURE 17.22 The cross-over or recombination process to generate new conformations.

sequence of numbers instead. However, the principle is the same.

Evolutionary programming is slightly different from genetic algorithms in that cross overs do not take place and mutations alone are used to generate the next generation of chromosomes. The first and second generation chromosomes then undergo a 'tournament' to see which represents the most stable conformations. This is done by comparing each chromosome against a randomly chosen number of opposing chromosomes. Chromosomes representing stable conformations will score more 'wins' in the tournament and be preserved. Those with poor scores are rejected.

Genetic and evolutionary programs are designed to find a stable conformation in a short time period, but due to their random nature, they cannot be guaranteed to find the global optimum. However, by carrying out several runs it is possible to find a variety of different stable conformations for a molecule. The method is best used with highly flexible molecules with more than eight rotatable molecules—molecules which would prove difficult to study using systematic searching.

17.9 Structure comparisons and overlays

Using molecular modelling, it is possible to compare the 3D structures of two or more molecules. For example, suppose we wish to compare the structures of **cocaine** and **procaine**. Both of these compounds have a local anaesthetic property, and structure–activity relationships indicate that the important pharmacophore for local anaesthesia is the presence of an amine, an ester and an aromatic ring. These functional groups are present in cocaine and procaine, but the pharmacophore also requires the functional groups to be in the same relative positions in space with respect to each other. Looking at the 2D structures of procaine and cocaine, it would be tempting to match up corresponding bonds as shown in Fig. 17.23, but this would place the nitrogen atoms one bond length apart in the overlay.

With molecular modelling, the important atoms of the structures can be matched up, in this case the nitrogens and the aromatic rings of both structures. The software then strives to find the best fit, resulting in the overlay shown in Fig. 17.24. Here the procaine molecule has been laid across the centre of the bicyclic system in cocaine, so that both the aromatic rings and nitrogen atoms overlap.

How does a software program know when a best fit has been achieved? This is done by calculating the root mean square distance (RMSD) between all the atom pairs which have to be matched up, and to find the relative orientation of the molecules where this value is a minimum. For example, in the overlay between cocaine and procaine, the pairs of atoms to be matched up are defined as the two nitrogens in each molecule and the corresponding aromatic carbons. The distance between the nitrogens in each molecule is measured, as are the distances between each of the corresponding aromatic carbons. The RMSD for all the atom pairs is then calculated. One of the structures is then moved in stages with respect to the other and the calculations repeated until a minimum value of RMSD is obtained, corresponding to the best fit.

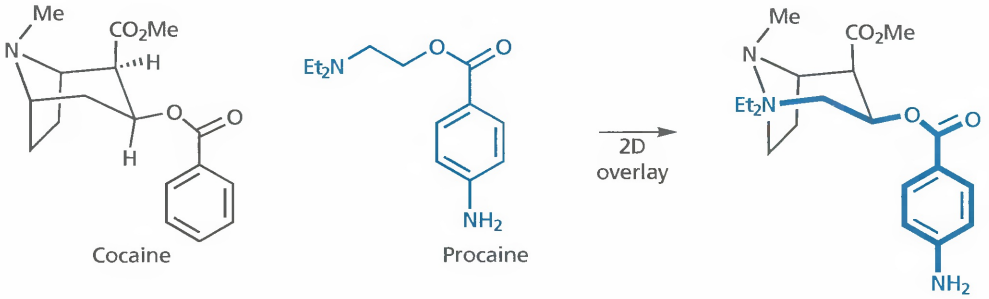


FIGURE 17.23 Two-dimensional overlay of cocaine and procaine.

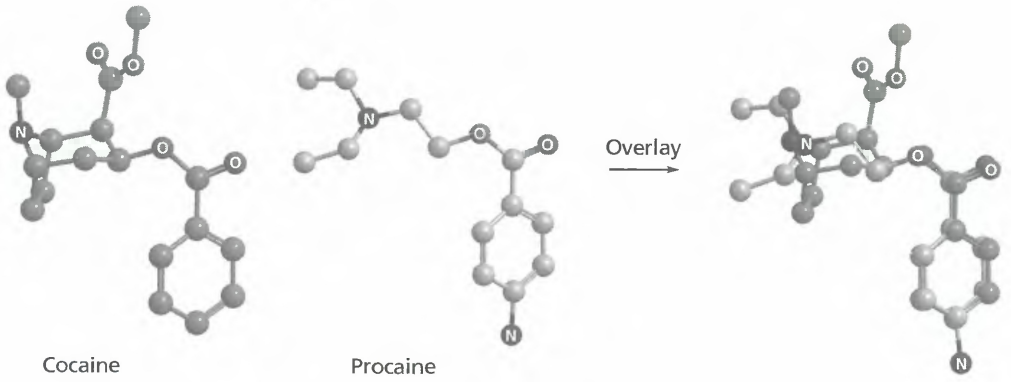


FIGURE 17.24 Overlay of cocaine and procaine using Chem 3D.

It is important to appreciate that the fitting process described above is carried out on a rigid basis; that is, the molecules are locked in the one conformation and no bond rotations are permitted. Therefore, it is important that each molecule is in the active conformation before carrying out the fitting process. If the active conformations are not known, it is possible to carry out overlays where one or both molecules can change conformations in order to get the best fit, but this would be more expensive on computer time.

Some modelling software programs have the capacity to automatically overlay two molecules without the operator having to define the centres or how they should be matched up. The program searches each molecule for what it considers to be important centres. These are centres which are normally involved in binding interactions (i.e. aromatic rings, hydrogen bond donors, hydrogen bond acceptors, positively charge centres, acidic centres, and basic centres). As far as an aromatic ring is concerned, the centre of the ring (the centroid) represents the whole ring. For hydrogen bond acceptors (X) or hydrogen bond donors (X–H), the heteroatom (X) is defined as the centre. To be precise, the centre for a hydrogen bond donor should be the hydrogen atom, but there is a large uncertainty as to where this atom is located because of bond rotation, and so the heteroatom is defined as the centre of an 'available volume' within which the hydrogen atom is located. Certain functional groups can be defined as being more than one type of centre. For example, the hydroxyl group is considered both as a hydrogen bond donor and a hydrogen bond acceptor centred on the oxygen. A primary amine is considered a hydrogen bond donor, hydrogen bond acceptor, basic centre, and positively charged centre because it could be protonated. Once the centres for each molecule have been identified, the program then strives to overlay them such that equivalent centres are matched up.

KEY POINTS

- Energy minimization produces the nearest stable conformation to the structure presented, and not necessarily the global conformation.
- Molecular dynamics can be carried out on a molecule to generate different conformations, which, on energy minimization, give a range of stable conformations. Alternatively, bonds can be rotated in a stepwise process to generate different conformations.
- Monte Carlo methods allow the random generation of stable conformations. The Metropolis method (simulated annealing) allows the identification of the most stable conformations and may lead to the identification of the global minimum.
- Genetic and evolutionary algorithms are used to generate different conformations and are designed to identify the most stable conformations. They may identify the global minimum but this cannot be guaranteed.
- Molecular modelling can be used to overlay two molecules in order to assess their similarity.

17.10 Identifying the active conformation

A problem frequently encountered in drug design is trying to decide what shape or conformation a molecule is in when it fits its target binding site—the active conformation. This is particularly true for simple flexible molecules which can adopt a large number of conformations. One might think that the most stable conformation will be the active conformation, as the molecule is most likely to be in that conformation. However, it is possible that a less stable conformation could be the active conformation. This is because the binding interactions with the target result in an energy gain which may compensate for the energy required to adopt that conformation.

17.10.1 X-ray crystallography

The easiest way of identifying an active conformation is to study the X-ray crystal structure of a target protein with its ligand (the drug) attached. The crystal structure

of the ligand itself can be obtained from the Cambridge Structural Database (CSD), and the crystal structure of protein-ligand complexes can be obtained from the Brookhaven National Laboratory Protein Data Bank (PDB) The protein–ligand complex can be downloaded and studied using molecular modelling software and the active conformation of the ligand identified. Not all proteins can be easily crystallized, however, and other methods of identifying active conformations may have to be used.

17.10.2 Comparison of rigid and non-rigid ligands

Identification of active conformations is made much easier if one of the active compounds is a rigid molecule which has only one possible conformation. The geometry of the pharmacophore (the important binding centres) can thus be determined for the rigid molecule. More flexible molecules can then be analysed to find a conformation which will place the important binding groups in the same relative geometry (Box 17.4).

BOX 17.4 Identification of an active conformation

The neuromuscular blocking agent **tubocurarine** is a fairly rigid structure where the two quaternary nitrogen atoms represent the pharmacophore. Molecular modelling allows the distance between these atoms to be measured as 11.527 Å (Fig. 1). **Decamethonium** also acts as a neuromuscular blocking agent, but it is an extremely flexible molecule, which means that a large number of conformations are possible. The most stable conformation is the extended one where the quaternary nitrogens are 14.004 Å apart. Using molecular dynamics, a variety of different conformations for decamethonium can be generated as described in section 17.8.2. One of these conformations has the quaternary nitrogens 11.375 Å apart – a possible candidate for the active conformation (Fig. 2).

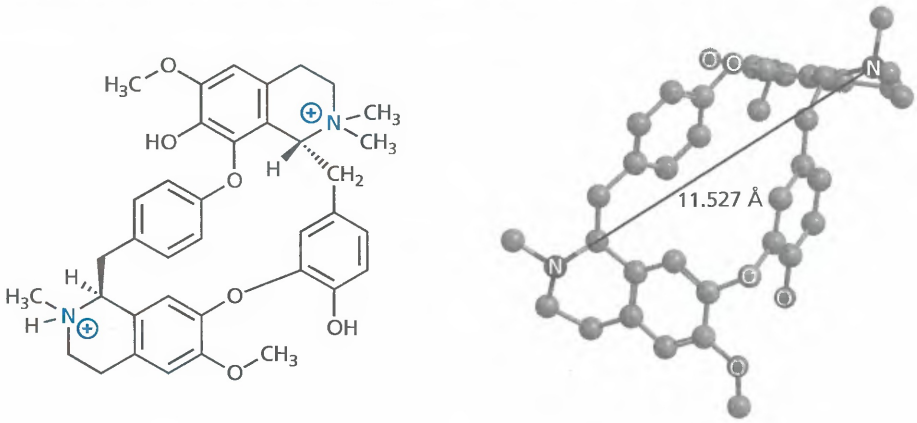


FIGURE 1 Computer generated model of tubocurarine (Chem 3D).

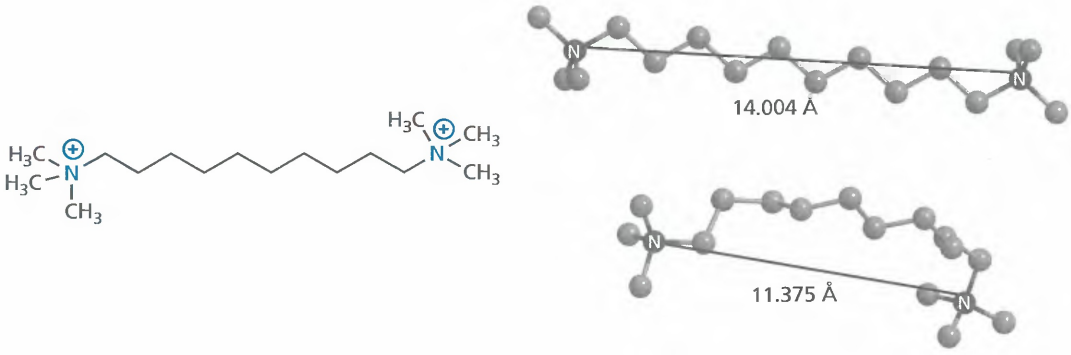


FIGURE 2 Computer generated conformations of decamethonium (Chem 3D).

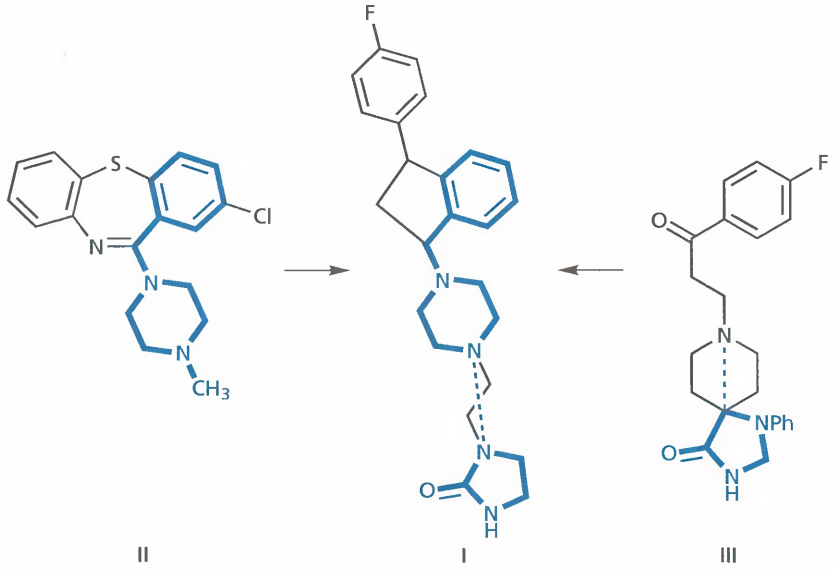


FIGURE 17.25 Defining the active conformation of serotonin antagonists. The dashed lines represent dummy bonds defining the relative positions of connected atoms.

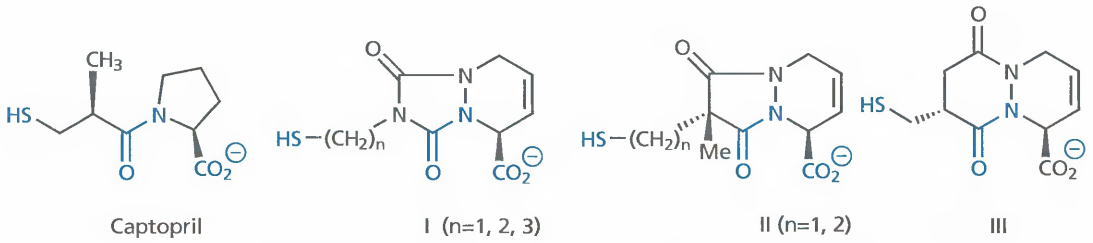


FIGURE 17.26 Captopril and rigid analogues (binding groups coloured).

If a fully rigid molecule is not available to act as a template, it may be possible to match up different structures which have an element of rigidity somewhere in their skeleton. For example, structures I–III in Fig. 17.25 are all antagonists of the $5HT_{2A}$ (serotonin) receptor. An active conformation for structure I could be proposed that matches the rigid moieties of structures II and III. Structure II gives the conformation for the top half of the molecule, and structure III gives the conformation of the bottom half.

Another method of finding active conformations is to consider all the reasonable conformations for a range of active compounds, and then to determine the common volume or space into which the various important binding groups can be placed in order to interact with the binding site. A study such as this was carried out in order to determine the active conformation of the anti-hypertensive agent **captopril** (Fig. 17.26). Because captopril is flexible, the exact 3D relationship of the important binding groups (i.e. the carboxyl, amide and thiol groups) in the active conformation is not known. There was also no X-ray crystallographic data available to reveal how captopril binds to its target binding site. To address this problem, a variety of rigid analogues (I–III) was synthesized, where the amide and the carboxyl group were fixed in space with respect to each other, but because of bond rotation the thiol group could access an area of space. The biological activity of these compounds was then measured to study how the relative orientation of the thiol group with respect to the other two groups affected biological activity.

The possible conformations for captopril arising from bond rotation around the two bonds shown in Fig. 17.27 were determined using molecular modelling, and a spatial map (A) was generated to show the possible regions in space that were accessible to the thiol group (Fig. 17.28).

Some of the conformations involved in this analysis are high-energy eclipsed conformations, and these were

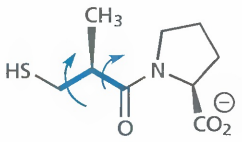


FIGURE 17.27 Bond rotations in captopril.

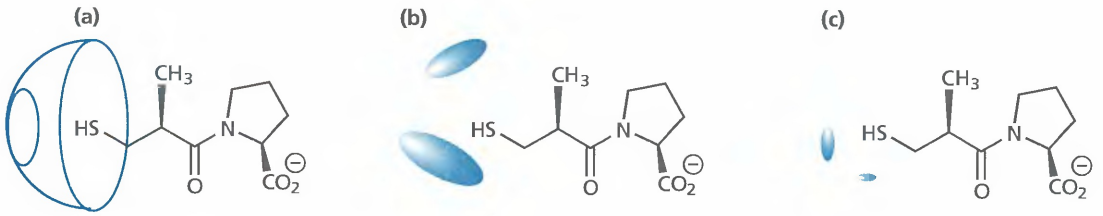


FIGURE 17.28 Generating spatial maps in conformational analysis. **(a)** All possible conformations, **(b)** stable conformations, and **(c)** after overlaps.

filtered out of the analysis by programming the software to reject conformations with steric energy greater than 200 kJ/mol. When this was done, the spatial map (B) showed that the thiol group was restricted to two main regions in space with respect to the other two binding groups.

A spatial map for one of the active rigid analogues was now generated in the same manner and compared with the one generated by captopril. The overlap between the maps was considered to be the most likely location for the thiol group. The process was then repeated for the other rigid analogues, further narrowing down the possible area that would be occupied by the thiol group. The study identified two 'hot spots' (C) for the thiol group. Conformations of captopril which placed the thiol group in those 'hot spots' were then considered to be likely active conformations.

17.11 3D pharmacophore identification

A 3D pharmacophore represents the relative position of important binding groups in space and disregards the molecular skeleton that holds them there. Thus, the 3D pharmacophore for a particular binding site should be common to all the various ligands which bind to it. Once the 3D pharmacophore has been identified, structures can be analysed to see whether they can adopt a stable conformation which will contain the required pharmacophore. If they do, and there are no steric clashes with the binding site, the structure should be active. There are several 3D chemical databases such as the CSD which can be searched for relevant structures.

17.11.1 X-ray crystallography

The crystal structure of a target protein with its ligand bound to the binding site can be used to identify the 3D pharmacophore. The protein-ligand complex can be downloaded on to the computer and studied to identify the bonding interactions that hold the ligand in the binding site. This is done by measuring the distances between likely binding groups in the drug and complementary binding groups in the binding site to see whether they are within bonding distance. Once the binding groups on the ligand have been identified, their positions can be mapped to produce the pharmacophore. The Brookhaven Protein Databank stores the crystal coordinates of proteins and other large macromolecules with and without bound ligands.

17.11.2 Structural comparison of active compounds

If the structure of the target is unknown, a 3D pharmacophore can be identified from the structures of a range of active compounds. Ideally, the active conformations and the important binding groups of the various compounds should be known. The molecules can then be overlaid as previously described in section 17.9 to ensure that the important binding groups are matched up as closely as possible. It would be rare for the binding groups to match up exactly, so an allowed region in space for each important binding group can be identified for the 3D pharmacophore.

17.11.3 Automatic identification of pharmacophores

It is possible to identify possible 3D pharmacophores for a range of active compounds using some software programs, even if the important binding groups are unknown or uncertain. First of all, the program identifies potential binding centres in a particular molecule. These are the hydrogen bond donors, hydrogen bond acceptors, aromatic rings, acidic groups, and basic groups. It is also possible to search for hydrophobic centres involving hydrocarbon skeletons of three or more carbon atoms. Here, the hydrophobic centre is calculated as the midpoint of the carbon atoms in question.

If dopamine was the structure being analysed, four important binding centres would be identified—the aromatic ring, both phenolic groups (hydrogen bond donors and acceptors) and the amine nitrogen (hydrogen bond donor, hydrogen bond acceptor, base, and a positively charged centre if protonated) (Fig. 17.29).

The program now identifies the various triangles which connect up the important centres. In the case of dopamine, there are apparently four such triangles. Each one is defined by the length of each side and the type of binding centres

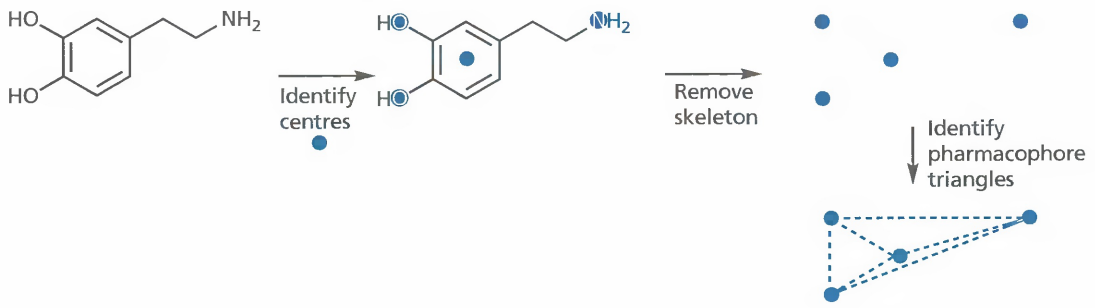


FIGURE 17.29 Pharmacophore ID in dopamine.

present, resulting in a set of pharmacophore triangles. Some of the points specified represent more than one type of binding centre, so this means that there will actually be more than four pharmacophore triangles. For example, if one of the points is a phenol then it represents a hydrogen bond donor or a hydrogen bond acceptor. Therefore, any triangle including this point must result in two pharmacophore triangles—one for the hydrogen bond donor and one for the hydrogen bond acceptor.

Of course, this analysis has only been carried out on one conformation of dopamine. The program is now used to generate a range of different conformations as described in section 17.8, and for each conformation another set of pharmacophore triangles is defined. Adding all these together gives the total number of possible pharmacophore triangles for dopamine in all the conformations created.

Another structure with dopamine-like activity is now analysed. Once all of its pharmacophore triangles have been determined, they are compared with those for dopamine and the pharmacophores that are common to both structures are identified. The process is then repeated for all the active compounds until pharmacophore triangles common to all the structures have been identified. These are then plotted on a 3D plot where the x-, y-, and z-axes correspond to the lengths of the three sides of each triangle. This produces a visual display which allows easy identification of distinct pharmacophores. Closely similar pharmacophores can be quickly spotted as they are clustered close together in specific regions of the plot.

For example, the three structures in Fig. 17.30 were analysed and found to have 38 common pharmacophores. When these were plotted, seven distinct groups of pharmacophore were identified (Fig. 17.31). Each

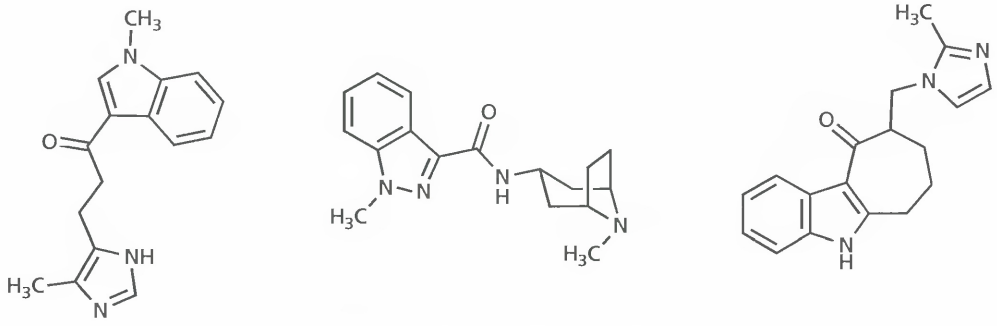


FIGURE 17.30 Test structures.

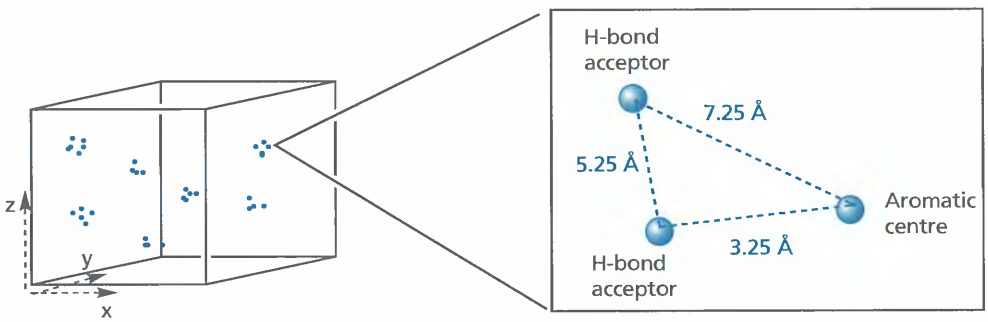


FIGURE 17.31 Pharmacophore plot.

pharmacophore present in the grid can be highlighted and revealed. For example, one of the possible active pharmacophores consists of two hydrogen bond acceptors and an aromatic centre. Note that it is advisable to begin this exercise with the most active compound and then proceed through the structures in order of activity.

The above analysis can be simplified enormously if certain groups are known to be essential for binding. The program can then be run such that only triangles containing these centres are included. For example, if the nitrogen atom of dopamine is known to be an essential bonding centre, then the triangles connecting the phenolic oxygens and the aromatic ring for each conformation can be omitted.

17.12 Docking procedures

17.12.1 Manual docking

Molecular modelling can be used to dock or fit a molecule into a model of its binding site. If the binding groups on the ligand and the binding site are known, they can be defined by the operator such that each binding group in the ligand is paired with its complementary group in the binding site. The ideal bonding distance for each potential interaction is then defined and the docking procedure is started. The program then moves the molecule around within the binding site to try and get the best fit as defined by the operator. In essence, the procedure is similar to the overlay or fitting process described in section 17.9, only this time the paired groups are not directly overlaid but fitted such that the groups are within preferred bonding distances of each other. Both the ligand and the protein remain in the same conformation throughout the process and so this is a rigid fit. Once a molecule has been successfully docked, fit optimization is carried out. This is essentially the same as energy minimization, but carried out on the ligand–target protein complex. Different conformations of the molecule can be docked in the same way and the interaction energies measured to identify which conformation fits the best.

17.12.2 Automatic docking

There are now a variety of docking programs that can automatically dock ligands into a binding site with the minimum of input from an operator. They have the advantage that they do not depend on any preconceived ideas that the operator may have on how a particular ligand should bind, and as a result they can reveal unexpected binding modes. They are also amenable to studying many different molecules automatically. Indeed, an important application of automatic docking programs is to carry out virtual screening of hundreds of different molecules with the aim of identifying new lead compounds that will interact with the target. Virtual screening can be seen as being complementary to biological screening, in that the former can identify the structures from a chemical 'library' that are most likely to bind to the target. These can then be given priority for biological screening, making the latter more efficient. For virtual screening to be effective, it has to use efficient algorithms which not only dock each molecule realistically, but also give an accurate 'score' of the relative binding energies of the molecules concerned. Moreover, for each molecule studied, the docking program is likely to generate several different orientations or binding modes. It is necessary to score all of these in order to identify the most likely binding mode in terms of how well it fits the space available and how many intermolecular interactions it can form with the binding site.

The calculations required for docking and scoring have to be rapid in order to process the number of molecules involved in a reasonable time period, but they also have to be accurate enough to give a good measure of relative binding energies. This is a difficult compromise to make since increasing the speed at which an algorithm operates involves assumptions or short cuts that inevitably reduce the accuracy of the calculation. As a result, this is an area of intense research interest in the development of new and improved docking programs. For reasons of space, it is not possible to go into the mathematical details of docking algorithms, and so this section focuses more on the general methods by which automatic docking can be carried out.

The simplest approach to automatic docking is to treat the ligand and the macromolecular target as rigid bodies. This is acceptable if the active conformation of the ligand is known, or if the ligand is a rigid cyclic structure. At the next level of complexity, the target is still considered as a rigid body, but the ligand is allowed to be flexible and can adopt different conformations. The most complex situation is where both the target and the ligand are considered to be flexible. This last situation is extremely expensive in terms of computer time, and most docking studies are carried out by assuming a rigid target.

17.12.3 Defining the molecular surface of a binding site

In order to carry out docking calculations, it is necessary to know the structure of the protein target and the nature of the binding site. This can be obtained from an X-ray crystal structure of the protein which can be downloaded

on to the computer. The amino acids lining the binding pocket can then be identified.

The next step is to define the molecular surface of the binding site. One could do this by defining each atom within the binding site by its van der Waals radius, but this results in an extensive surface area, much of which would be inaccessible to a ligand (Fig. 17.32).

A simpler molecular surface can be defined by identifying the parts of the van der Waals surface that are accessible to a solvent molecule. In practice, a probe sphere of radius 1.4–1.5A is used to represent a water molecule, and this is 'rolled over' the surface of the binding site (Fig. 17.33). Convex surfaces shown in dark blue are where the probe sphere makes contact with the van der Waals surfaces of a particular atom. Concave surfaces shown in light blue are known as re-entrants and represent how far the probe atom can access the space between the atoms of the binding site. In this area, the probe is in contact with two or three atoms. This kind of molecular surface is also referred to as a **Connolly surface**. The surface is actually represented by a regular distribution of points or dots, and the crucial ones for docking are those on the convex surfaces. Each one of these has a vector associated with it which points into the binding site. The direction of the vector corresponds to the normal of the surface at that point, and so it is a mathematical indication of curvature.

17.12.4 Rigid docking by shape complementarity

The first problem with any docking program is how to position the ligand within the binding site. If you or I were handed real models of the target and the ligand, we would consider the space available in the binding site, eye up the ligand and judge how we could place it into the binding site before we actually do it. In other words, humans have a spatial awareness which includes the ability to assess the shape of an empty space. This does not come naturally to computers and the empty space of a binding site has to be defined in a way that a computer program can understand before ligands can be inserted.

The **DOCK program** was one of the earliest programs to tackle this problem. The Connolly surface is

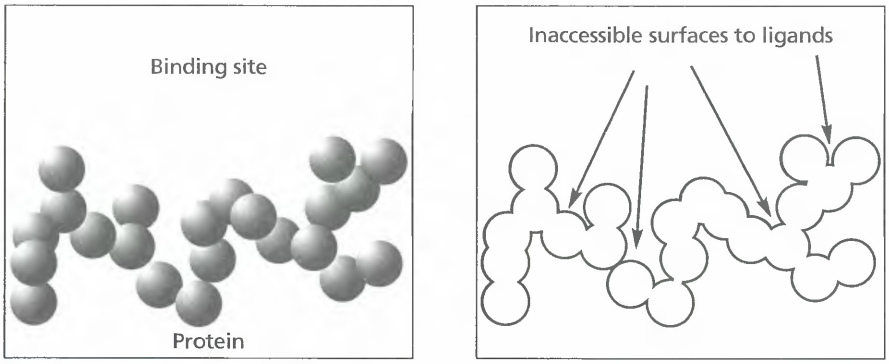


FIGURE 17.32 Defining the surface of the binding site by atoms and van der Waals surfaces.

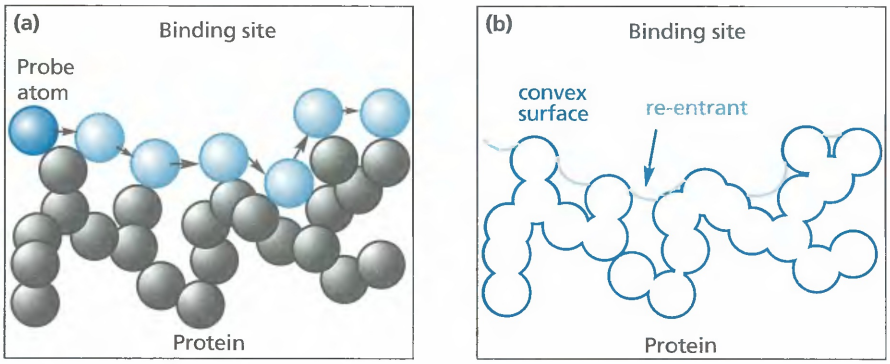


FIGURE 17.33 **(a)** Defining the Connolly surface of a binding site with a probe atom. **(b)** The Connolly surface shown in blue.

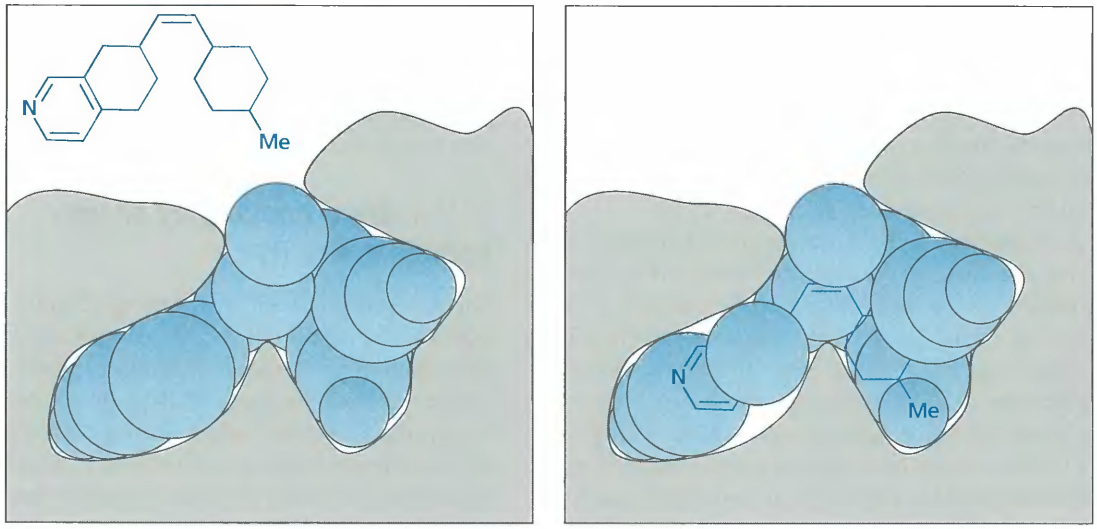


FIGURE 17.34 The DOCK program.

first defined, then the empty space of the binding site is defined by identifying a collection of differently sized spheres which will fill up the space available and give a 'negative image' of the binding site (Fig. 17.34). This is achieved as follows.

For each dot representing the molecular surface, spheres are constructed that touch that dot plus one other dot on the molecular surface. Therefore, if there are n dots representing the molecular surface, n-1 spheres will be created at *each* of the dots. This represents a massive number of spheres and so it is necessary to whittle these down. The number of spheres can be significantly reduced as follows:

- For each dot on the molecular surface, the sphere of smallest radius touching it is chosen. This ensures that none of the spheres chosen intersects the molecular surface.
- There are several dots associated with the surface of a particular atom and each of these now has one sphere associated with it. The next filtering process is to select the sphere with the largest radius. Once this has been completed, the number of spheres left is the same as the number of atoms lining the binding site. Spheres are allowed to overlap and the centre of each sphere accurately defines a unique position of 3D space within the binding site.

Each sphere representing the binding site can be considered as a 'pseudoatom' and so it is now possible to carry out an overlay operation as described in section 17.9 where ligand atoms are matched with binding site spheres, then overlaid. However, how does the programme decide which ligand atom and receptor spheres should be matched? One could try out every possible combination, but this would take up far too much computer time. Instead, a systematic matching operation takes place called **distance matching** or **clique searching**. First, the distances between each of the ligand atoms are measured. This is repeated for all of the receptor spheres. These distances are then used to identify which ligand atoms and receptor spheres can be matched up. The operation takes place like this. A graph is prepared where each ligand atom (1, 2, 3, . . .) is matched to each of the receptor spheres (A, B, C, . .) to give a list of paired atom/spheres (1A, 1B, 1C. . ., 2A, 2B, 2C . . . . 3A, 3B, 3C . . . etc). The next stage is to identify whether two of these pairs are compatible; for example is the pairing 1A possible at the same time as the pairing 2C? This is done by comparing the distance between the ligand atoms 1 and 2, with the distance between the receptor spheres A and C. If the distances are similar, then they are compatible. This process is now repeated for further pairings to see if they are compatible with those already identified. The minimum number of pairings required for an acceptable docking is four. The whole procedure is repeated systematically for each ligand atom to find a variety of matches which will eventually lead to different docking modes.

As an example, consider a ligand represented by atoms 1–10, and a binding site represented by spheres A–G (Fig. 17.35). The atom/sphere pairs 1A, 6F, 7G and 10E would be identified as compatible for the docking operation since the distances between the specified ligand atoms match the distances between the specified receptor spheres.

Only once this procedure has been carried out will the actual docking process take place. Docking then involves an overlay where ligand atoms are fitted onto their paired

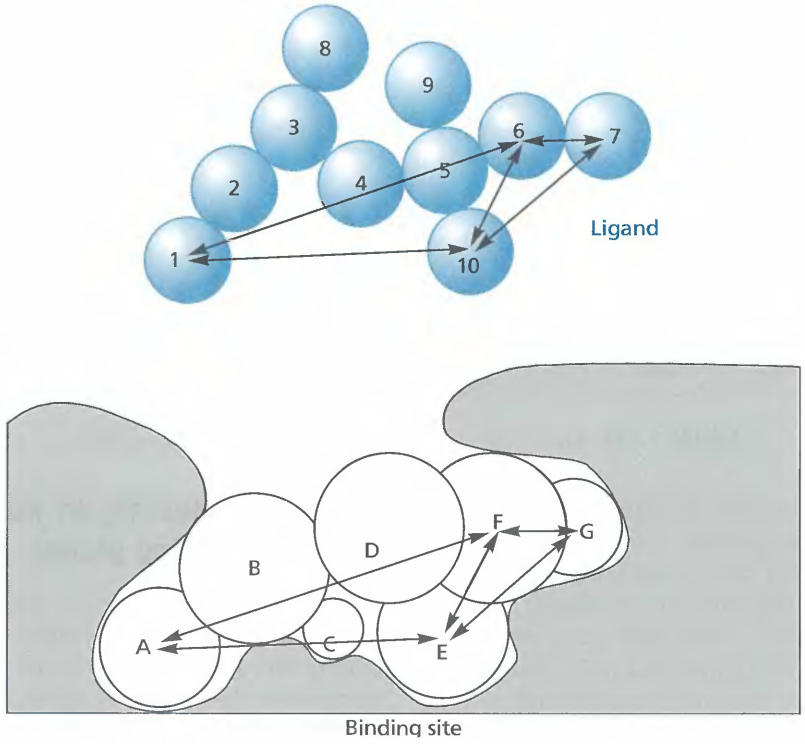


FIGURE 17.35 Comparing distances between ligand atoms and distances between binding site spheres.

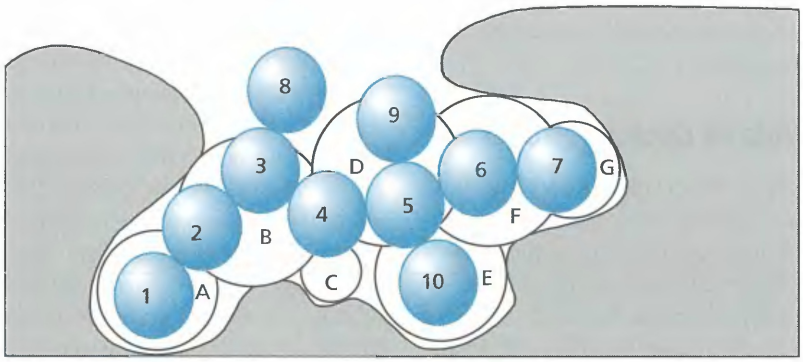


FIGURE 17.36 Docking by overlaying the atom-sphere pairings of 1A, 6F, 7G and 10E.

receptor spheres as described in section 17.9. For example, in Fig. 17.36, ligand atoms 1, 6, 7, and 10 are matched to receptors spheres A, F, G and E respectively (Fig. 17.36). This process is repeated for all the other possible matches to give a number of docking or binding modes.

Note that this docking procedure is carried out purely in terms of steric complementarity (i.e. whether selected ligand spheres can match up with selected receptor atoms). It takes no account of possible binding interactions, either favourable or unfavourable. Moreover, since selected ligand atoms are matched up with selected receptor spheres, it is perfectly possibly that some of the binding modes obtained are impossible. For example, a ligand atom not used in the matching operation might be placed in the same space as an atom lining the binding site (Fig. 17.37). Therefore, a filtering process has to be included in the program to remove any such unacceptable binding modes.

If the binding mode is acceptable, an optimization process is carried out which 'fine tunes' the position of the ligand in the binding site. This minimizes unfavourable steric interactions and optimizes intermolecular interactions between the ligand and the binding site. The binding energy of the ligand is now measured and a score is given for that binding mode.

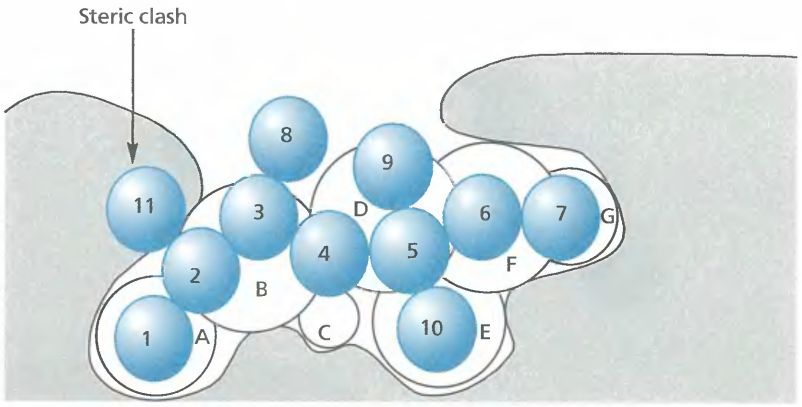


FIGURE 17.37 An unacceptable binding mode due to a steric clash.

This is repeated for all the possible matches and binding modes. The binding modes with the highest scores are then stored so that they can be analysed further by the operator. In the original version of DOCK, this scoring operation took into account only steric interactions and hydrogen bond interactions, but many other factors have an influence on receptor-ligand binding such as other types of intermolecular interactions, desolvation, the difference in energy between a ligand's different conformations, and the decrease in entropy resulting from a flexible molecule being bound in a fixed conformation. Later versions of DOCK have tackled these issues, as have other docking programs.

17.12.5 Use of Grids in docking programs

A major step forward in the development of docking programs was the use of grids to pre-calculate the binding interactions at different positions within the binding site (section 17.7.5). These values are stored in 'look-up' tables and automatically accessed for each atom of the ligand based on its particular position within the grid. This means that binding energies can be speedily obtained by adding up the relevant table entries rather than measuring the interaction between each ligand and the binding site—a process which would take far longer. The first docking programs to implement the use of grids in this way were **AutoDock** and a revised version of **DOCK**, and they are now commonly used in many other programs.

17.12.6 Rigid docking by matching hydrogen bonding groups

The docking process described in section 17.12.4 is based on whether a ligand has the right shape to fit into the binding site, and takes no account of possible binding interactions. This is ideal for ligands that take up most of the space available in the binding site, but is less satisfactory for ligands which are small in comparison to the size of the binding site.

An alternative method of docking is to use the same 'clique technique' described in section 17.12.4, but this time to match up hydrogen bonding groups in the binding site with complementary hydrogen bonding groups present on the ligand. There are two important factors to take into account. First, a hydrogen bonding group on the ligand must be the correct distance from a hydrogen bonding group in the binding site. Second, the two groups concerned must have the correct orientation with respect to each other (section 1.3.2). It is therefore necessary to identify positions in space within the binding site where ligand atoms can be positioned to satisfy these criteria. These positions are defined by interaction points as follows.

First, a sphere is created around each hydrogen bonding group in the binding site (Fig. 17.38). The surface of the sphere represents the optimum distance at which a complementary group on the ligand should be placed in order to form a good hydrogen bond. A series of uniformly spaced points is placed over the surface of the

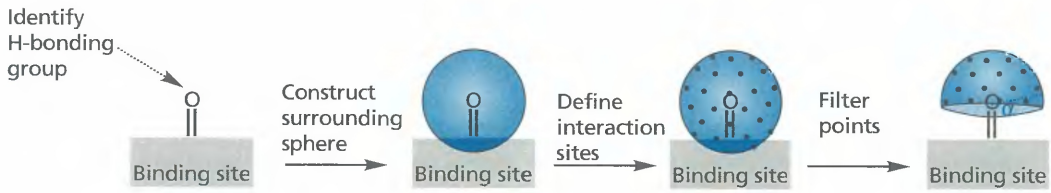


FIGURE 17.38 Identifying interaction sites for hydrogen bonding groups in the binding site.

sphere to define the surface. These are the interaction points onto which complementary binding groups on the ligand will be positioned during the docking process. However, not all of the points are feasible positions for a good hydrogen bonding interaction, and so a filtering process takes place which:

- removes the points that are not in the binding site
- removes the points which would not allow a bonding angle (α) greater than 90°.

The interaction points that survive this filtering procedure are now used as 'targets' for the matching operation with suitable ligand atoms.

This method is used in the **Directed Dock** algorithm alongside the matching algorithm based on shape complementarity. This means that it is possible to carry out a docking which takes into account both hydrogen bonding interactions and shape complementarity.

17.12.7 Rigid docking of flexible ligands: the FLOG program

One of the major drawbacks of rigid docking experiments is that they may fail to give a satisfactory answer for flexible ligands. Such ligands can form a variety of different conformations, and unless one knows the active conformation, it is a matter of chance whether the conformation chosen for the docking experiment is the ideal conformation for docking or not. One way round this is to dock as many different conformations of the ligand as possible in order to get the best result. **FLOG** (Flexible Ligands Orientated on Grid) is a docking program that generates conformational libraries called **Flexibases,** which contain 10–20 conformations for each ligand studied. However, there is still a chance that the correct conformation will not be tested, especially for very flexible ligands. The more flexible the ligand, the more conformations that are possible and this can lead to a **'conformational explosion'**. In other words, the number of possible conformations increases exponentially with the number of rotatable bonds present.

17.12.8 Docking of flexible ligands: anchor and grow programs

Various program have been written to allow the generation of different ligand conformations as part of the docking process. A popular method is to fragment the ligand, identify a rigid anchor fragment which can be docked, then reconstruct or grow the molecule back onto the anchor. The following are examples of such programs.

Directed Dock and Dock 4.0

Directed Dock and DOCK 4.0 use a method where the algorithm identifies the rotatable bonds that are present in a ligand, allowing the identification of rigid and flexible regions. The molecule is then split into molecular components or fragments (Fig. 17.39).

The most rigid fragment is defined as the anchor and is docked by shape complementarity (section 17.12.4). The segments representing the flexible parts of the molecule are then added sequentially to the anchor. As each segment is added, torsion angles are varied in a systematic fashion and this inevitably increases the number of partially built structures (constructs). The number of constructs would go through the roof if this was to continue and so once each segment is added, there is a pruning process which selects a limited number of constructs based on how well they bind and also on how different they are in structure.

The segments are added in 'layers' working outwards. Thus, all the segments in layer 1 are added sequentially before the segments in layer 2 (Fig. 17.39). At each stage

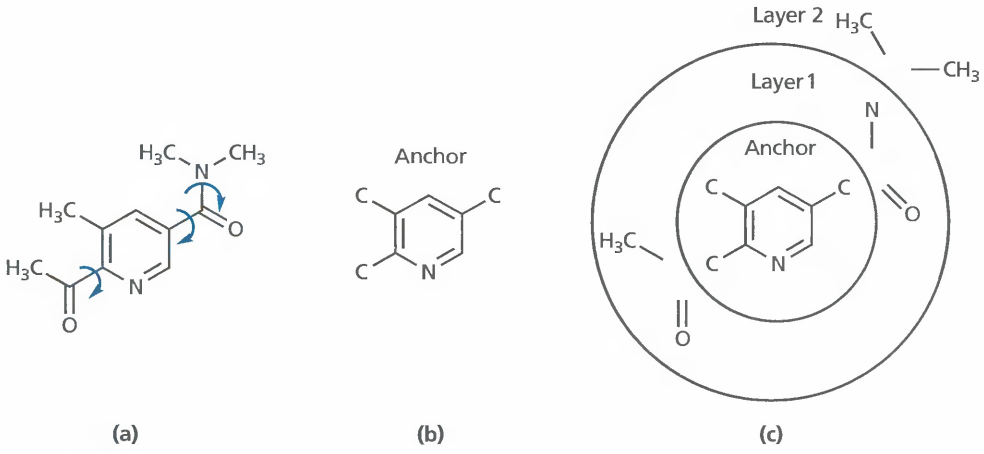


FIGURE 17.39 Anchor and grow algorithm. **(a)** Identify rotatable bonds. **(b)** Identify and dock a rigid anchor. **(c)** Add molecular fragments in layers.

of the process, energy minimization of the construct is carried out to relieve any strain arising from the construction process.

FlexX

FlexX is a software progam that also uses the anchor and grow method, but here the anchor is docked according to chemical complementarity—in other words docking is determined by the intermolecular interactions that can be formed between the anchor and the binding site. Docking the anchor by chemical complementarity rather then steric complementarity has the advantage that it cuts down the number of possible binding orientations for the anchor.

An interaction surface consisting of interaction points is built round each potential binding group in the binding site (Fig 17. 40) (see also section 17.12.6). A matching process now takes place which matches atoms on the anchor to interaction points in the binding site. The distances between atoms on the anchor must match the distance between interaction points in the binding site. This is the same procedure that is carried out in section 17.12.4, but there is the added requirement that the anchor atom and the corresponding interaction point must have binding compatibility. Docking requires the identification of three matched pairs of anchor atoms/interaction points. This corresponds to identifying complementary pharmacophore triangles for the anchor and the binding site.

The matching process is thus brought down to a comparison of pharmacophore triangles for the ligand and pharmacophore triangles for the binding site. For a match to occur, a triangle for the ligand must have roughly the same dimensions as a triangle for the binding site, but also the corners of the triangle must have binding compatibility.

The docking is now carried out such that anchor atoms are overlaid with their matched interaction point in the binding site (Fig. 17.41).

The procedure ensures that the angle requirements for hydrogen bonding are fine with respect to the interaction points in the binding site, but angles with respect to the anchor atoms now have to be checked. This is possible since the program builds a set of interaction points round the anchor atoms prior to docking. For example, consider the case where docking matches up an N-H group on the anchor with a carbonyl group in the binding site (Fig. 17.42). The docking procedure fits the hydrogen of the N-H group onto an interaction point round the carbonyl group, and so the hydrogen is at the correct orientation and distance with respect to the oxygen. The program now checks to see whether the oxygen is placed on an interaction point round the hydrogen. If so, the oxygen is at the correct distance and

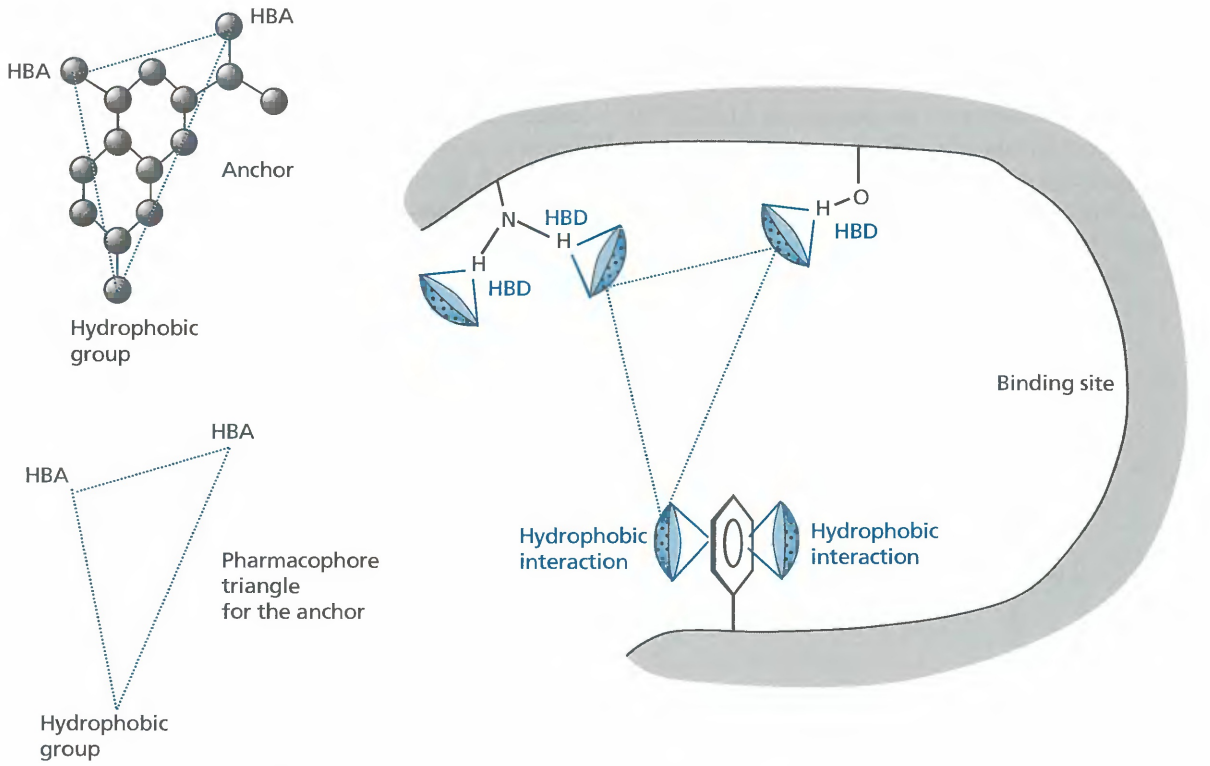


FIGURE 17.40 Docking based on the matching up of complementary pharmacophoric triangles on the anchor and the binding site (HBD = hydrogen bond donor, HBA = hydrogen bond acceptor).

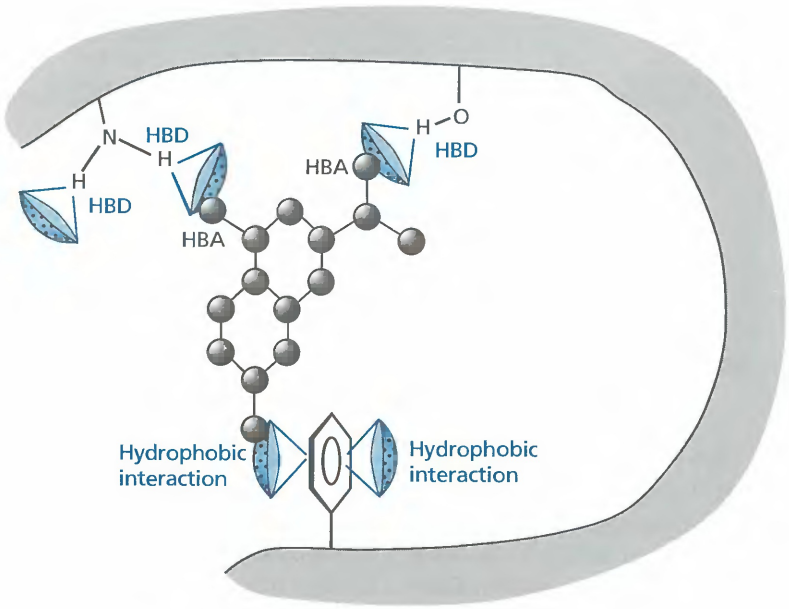


FIGURE 17.41 Docking of the anchor (HBD = hydrogen bond donor, HBA = hydrogen bond acceptor).

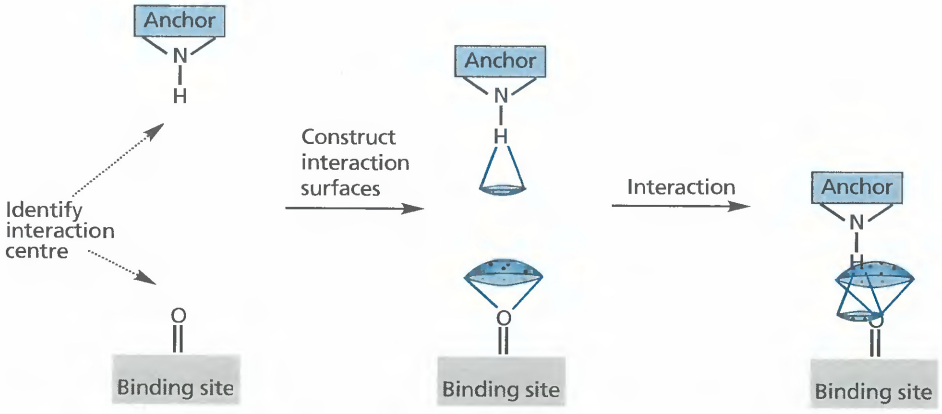


FIGURE 17.42 Assessing whether groups are at the correct distance and orientation with respect to each other for a good hydrogen bond interaction.

orientation with respect to the hydrogen, and a good interaction exists.

There will be several docking solutions for the anchor. These are 'clustered' and a representative binding orientation is taken from each cluster. The remainder of the molecule is than added. Fragments are attached to the anchor by rejoining the linkers at a discrete set of torsion angles.

One problem associated with the program is that the anchor has to be chosen manually and this becomes an issue if one wants to do the automatic docking of a series of structures in a database. Another problem is the vast number of different pharmacophore triangles that could be constructed to represent the binding site. One has to remember that each binding group has an interaction surface represented by numerous interaction points, and so the number of triangles which could be built is enormous. There are ways in which the docking algorithm can store the equivalent information in a more efficient manner, but the details of that are beyond the scope of this text.

The Hammerhead program

The Hammerhead program is another anchor and grow procedure which was designed to carry out docking studies on a large number of compounds. For example, it has been used to study databases of 10 000–100 000 small molecules in a few days.

Probes are placed into the binding site in order to identify the optimum locations for particular binding

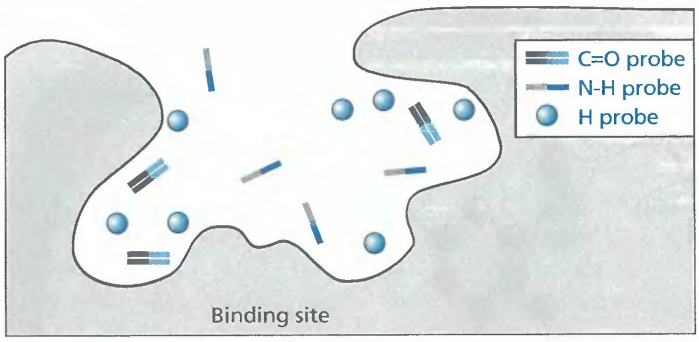


FIGURE 17.43 Positioning probes into a binding site to identify binding interactions.

interactions (Fig. 17.43). The probes used are hydrogen atoms, as well as C=O and N–H fragments. Each of the probes can be scored as high scoring or low scoring, based on the number of hydrogen bonds it can form or on favourable hydrophobic ring face contacts. Once the probes have been positioned, they act as the targets for the docking procedure. The same kind of matching operations described previously are carried out to match atoms of a molecular fragment with probes, and there is a requirement that docking must involve at least one of the high scoring probes. Both steric and chemical complementarity is used in the matching process and once the matches have been identified, the docking operation is carried out.

As far as the ligand is concerned, it is split into different fragments, each of which will have a limited number of rotatable bonds. All the fragments that are formed contain an atom or bond that is shared with another fragment (Fig. 17.44).

For each fragment, a number of conformations are generated. The fragments are docked and scored. Fragments that are particularly high scoring are defined as 'heads' and act as anchors. The remaining fragments are defined as 'tails'. The reconstruction phase is carried out for each fragment that has been identified as a potential anchor. Fragments that have been identified as 'tails' are then docked one at a time into the area around the anchor head. The first fragment shares an atom or a bond with the anchor and is docked such that it is aligned both to the relevant atom or bond on the anchor, as well as to the pocket probes. The two fragments are then merged ('chaining') by overlaying the shared atoms or bonds (Figs. 17.45–46). This involves the tail fragment being moved to the anchor rather than the other way round (i.e. the anchor remains fixed).

An optimization of the construct is carried out after each stage in order to enhance binding interactions with the binding site, and to remove any unfavourable steric interactions that might be present.

This method has advantages over procedures that manually choose a single anchor. Anchors are chosen automatically without any bias from the operator. Moreover, several different anchors are possible, all of which can be investigated. This is important since it has been observed that the anchor with the highest binding score does not necessarily produce the best docking mode for the final ligand.

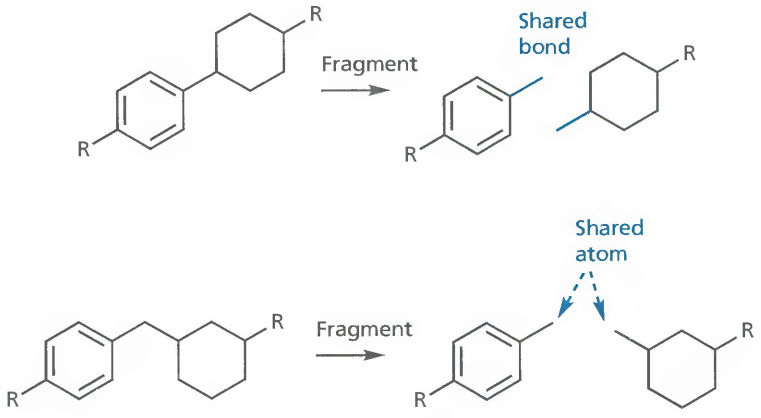


FIGURE 17.44 Fragmenting the ligand.

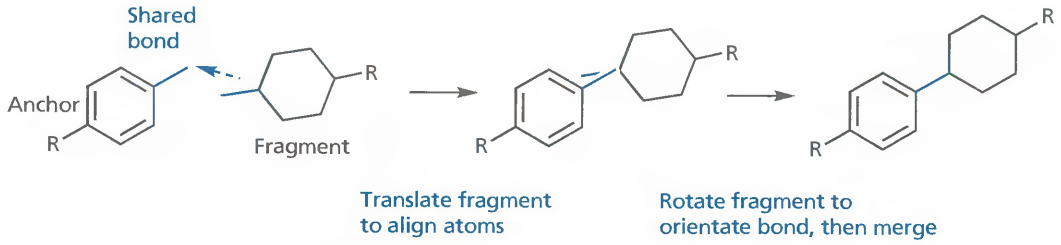


FIGURE 17.45 Merging fragments with a shared bond.

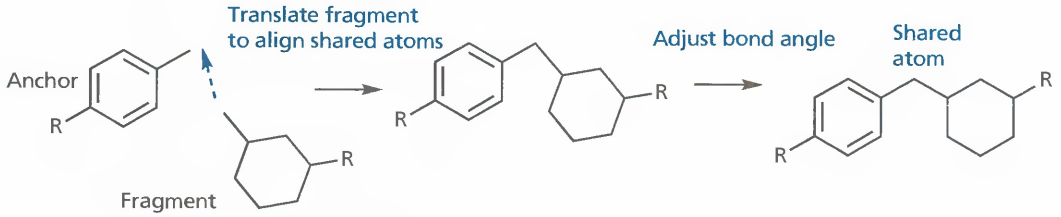


FIGURE 17.46 Merging fragments with a shared atom.

17.12.9 Docking of flexible ligands: simulated annealing and genetic algorithms

A number of docking programs use simulated annealing or genetic algorithms which incorporate a conformational search as part of the docking process. These methods are viable for the docking of flexible ligands, but compared to previous methods, they are slower and computationally more expensive.

The Metropolis method (or simulated annealing) involves the use of Monte Carlo algorithms and was described in section 17.8.4 as a method of conformational analysis. The same principles are involved in docking studies. The intact ligand is placed randomly in the space close to the binding site. Monte Carlo algorithms are then used to generate different conformations as described in section 17.8.4, but the whole molecule is also translated and rotated such that it 'tumbles' within the binding site. Different conformations are therefore generated at different positions and orientations within the binding site, and the binding energy of each structure is measured as it is formed and compared with the previous structure. Simulated annealing is carried out in order to identify the best binding modes. The principles of simulated annealing were explained in section 17.8.4 and hold true here—the only difference being that it is the binding energy of different binding modes which is measured rather than the steric energy of different conformations.

One of the reasons that the procedure is slower and more computationally expensive is the need to measure binding energies for each structure and binding mode as it is formed. This can be speeded up by using 'look-up tables' which are initially prepared by the grid and probe method (section 17.7.5).

A variety of docking programs use Monte Carlo algorithms for docking, including AutoDock, MCDOCK, Prodock and PRO-LEADS. A disadvantage of this approach is that the quality of results often depends on how the initial structure is placed in the binding site. Some research groups have used a combination of programs to address this problem. For example, DOCK can be used to identify binding modes for a specific ligand conformation based on a rigid fit and steric complementarity. Each of the binding modes identified can then be used as the starting structure for a Monte Carlo based docking program which generates different conformations and orientations in that area of the binding site.

Programs using evolutionary and genetic algorithms have also been used in docking studies. The principles of these programs were described earlier in section 17.8.4 as a method of generating different conformations. The same principles hold for docking. However, chromosomes are set up which not only determine the conformation of a molecule, but also its position and orientation within the binding site. Mutations and cross-over procedures not only change the conformation, they reorientate the molecule through translation and/or rotation of the whole molecule. Selection of the best docking modes is based on how well each molecule interacts with the binding site. Examples of programs that use genetic algorithms include AutoDock, GOLD and later versions of DOCK.

17.13 Automated screening of databases for lead compounds

The automated docking procedures described in section 17.12.2 can be used to screen a variety of different 3D structures to see whether they fit the binding site of a particular target (**electronic screening** or **database mining**). This is useful for a pharmaceutical company wishing to screen its own or other chemical stocks (libraries) for suitable lead compounds.

Screening of databases can also be done purely by searching for suitable pharmacophores. The process is speeded up by a quick filter which eliminates any structure that does not contain the necessary centres. The operator has the ability to vary the tolerances involved in the search in order to find pharmacophores which nearly match the desired pharmacophores.

17.14 Protein mapping

Drug design is made easier if the structure of the target protein and its binding site are known. The best way of obtaining this information is from X-crystallography of protein crystals, preferably with a ligand bound to the binding site. Unfortunately, not all proteins are easily crystallized (e.g. membrane proteins). In cases like this, model proteins and binding sites may be constructed to aid the drug design process.

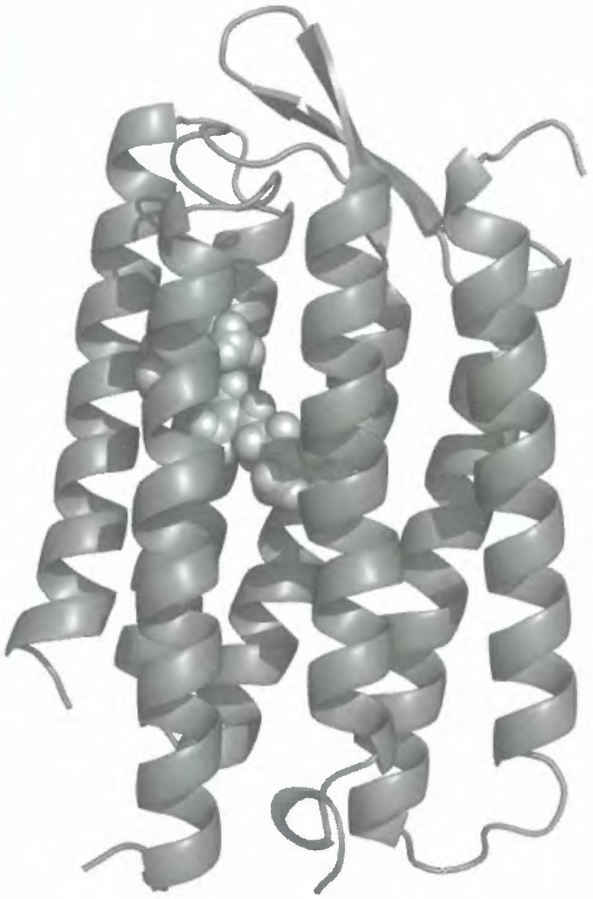


FIGURE 17.47 Bacteriorhodopsin with bound retinal.

17.14.1 Constructing a model protein: homology modelling

A model of a protein can be created using molecular modelling if the primary amino acid sequence is known and the X-ray structure of a related protein has been determined. Of historical interest in this respect is the protein **bacteriorhodopsin** which has been crystallized and its structure determined by X-ray crystallography (Fig. 17.47). Bacteriorhodopsin is structurally similar to **G-protein coupled receptors** which contain seven transmembrane helices (section 4.7). Many of the important receptors in medicinal chemistry belong to this family of proteins and so the structure of bacteriorhodopsin was used as a template to construct what is termed as **homology models** of these membrane bound receptors. By identifying the primary amino acid sequence of the target receptor and looking for suitable stretches of hydrophobic amino acids, it is possible to identify the seven transmembrane helices and then to use bacteriorhodopsin as a template in order to construct the helices in a similar position relative to each other. The linking loops can then be modelled in to give the total 3D structure. Unfortunately, bacteriorhodopsin is not a G-protein-coupled receptor and so it is not the ideal template for constructing these model receptors. In 2000, the crystal structure of the bovine receptor **rhodopsin** was successfully determined by X-ray crystallography. This *is* a G-protein-coupled receptor and provides a better template for the construction of more accurate receptor models. More recently in 2007, human β_2-adrenoceptor was crystallized in a lipid environment, with an inverse agonist bound to the binding site. An antibody fragment was bound to one of the intracellular loops in order to stabilize the structure. The X-ray crystal structure reveals the structure of the transmembrane helices and the intracellular regions, but the structure of the extracellular regions and the ligand binding site are still to be fully determined.

If a new protein has been discovered, its primary structure is first determined. Suitable software is then used to compare its primary sequence with the primary sequences of other proteins in order to find a closely related protein. This involves comparing the sequences with respect to conserved amino acids, hydrophobic regions and secondary structure. Once a reference

protein of similar structure has been identified, it is used as a template in order to build the peptide backbone of the new protein. First of all, regions which are similar in the new protein and the template protein are identified. The backbone for the new protein is constructed to match the corresponding region in the template protein. This leaves connecting regions whose structure cannot be determined from the template. A suitable conformation for these intervening sequences might be found by searching the protein databases for a similar sequence in another protein. Alternatively, a loop may be generated to connect two known regions. Once the backbone has been constructed, the side chains are added in energetically favourable conformations. Energy minimization is carried out and the structure is refined with molecular dynamics in the absence and presence of ligand. Once the model has been constructed, it is tested experimentally. For example, the model would indicate that certain amino acids might be important in the binding site. These could then be mutated to see if this has an effect on ligand binding. Studies such as this have identified amino acids which are important in binding neurotransmitters in a range of G-protein-coupled receptors Fig. 17.48).

This study shows interesting similarities and differences between the four receptors studied. For example, all four binding sites interact with ligands having a charged nitrogen group and contain a hydrophobic pocket lined with aromatic residues to receive it. There are several conserved aromatic residues in this pocket at positions 307, 613 and 616. An aspartate residue at position 311 is also present in all cases and is capable of forming an ionic interaction with the charged nitrogen on the ligand.

There are also significant differences between the binding sites which account for the different ligand selectivities. For example, the amino acids at positions 505 and 508 in the catecholamine receptors are serine, whereas the corresponding amino acids in the cholinergic receptors are alanine. The amino acid at position 617 in the catecholamine receptors is phenylalanine, allowing an interaction with the aromatic portion of the catecholamines, whereas in the cholinergic receptor this amino acid is asparagine, allowing a hydrogen bonding interaction with the ester group of acetylcholine.

17.14.2 Constructing a binding site: hypothetical pseudoreceptors

Rather than construct a complete model protein, it is possible to use molecular modelling to design a model binding site based on the structures of the compounds which bind to it. In order to do this effectively, a range of

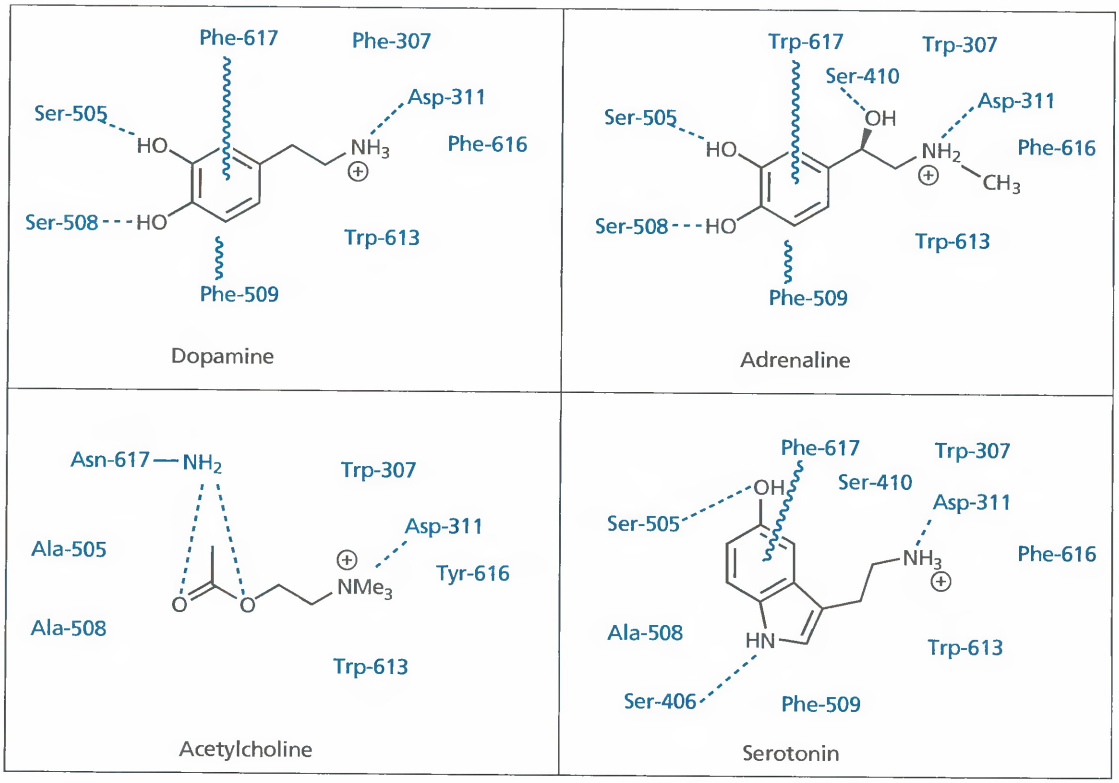


FIGURE 17.48 Important amino acids in various binding sites. (The numbering indicates the position of each amino acid on the seven possible helices of a G-protein-coupled receptor. For example, 311 indicates position 11 on helix 3.)

structurally different compounds with a range of activities is chosen. The active conformations are identified as far as possible and a 3D pharmacophore identified as described previously. The molecules are then aligned with each other such that their pharmacophores are matched. Each molecule is then placed in a potential energy grid and different probes are placed at each grid point in turn to measure interaction energies between the molecule and the probe atom (section 17.7.5). An aromatic CH probe is used to measure hydrophobic interactions, and an aliphatic OH probe is used to measure polar interactions. The interactions are then displayed by isoenergy contours (typically -6.0 kJ mol^{-1} for hydrophobic interactions and -17.0 kJ mol^{-1} for polar interactions).

The fields for all the molecules in the study can then be compared to identify common fields. Once these have been identified, suitable amino acids can be positioned to allow the required interaction. For example, an aspartate residue could be used to allow an ionic interaction, and amino acids such as a phenylalanine, tryptophan, isoleucine, leucine or valine could be used for a hydrophobic interaction (Box 17.5).

BOX 17.5 Constructing a receptor map

A range of structures including **altanserin** and **ketanserin** was used to construct a model receptor binding site for the 5-HT$_{2a}$ receptor. Taking ketanserin as the representative structure for these compounds, various hydrogen bonding, ionic bonding and hydrophobic bonding interactions were identified. Structure–activity relationships (SAR) were then used to identify whether any of these proposed interactions were important or not. In this case, SAR indicated that the two carbonyl groups were not important, and so the hydrogen bonding regions derived from these groups probably do not exist in the receptor binding site. Suitable amino acids can now be placed in the relevant positions. The choice of which amino acids should be used is helped by knowing the amino acid sequence of the target protein, and the structure of a comparable protein. The 5-HT$_{2a}$ receptor belongs to a superfamily of proteins that includes bacteriorhodopsin, whose structure is known. Allying this information with the primary amino acid sequence of the receptor led to the choice of amino acids shown in Fig. 1.

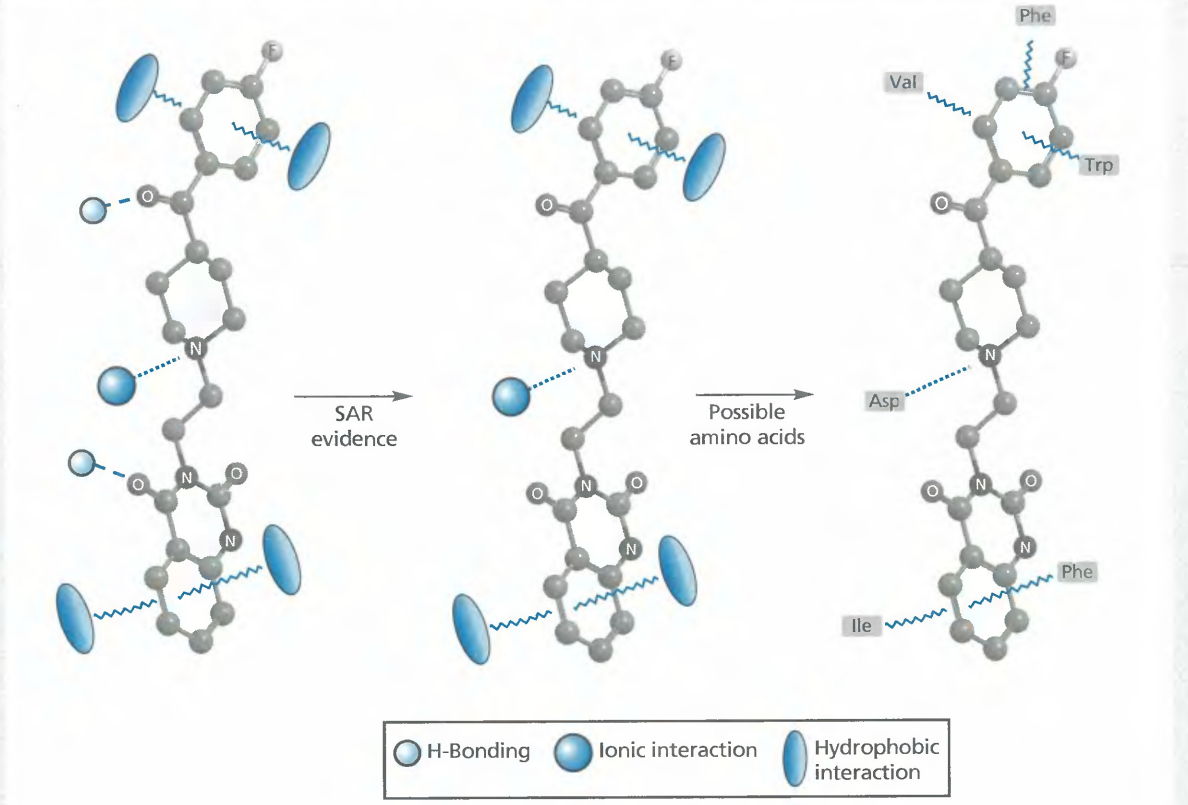


FIGURE 1 Receptor map for the 5-HT$_{2a}$ receptor using ketanserin as representative ligand.

Once built, known compounds can be docked to the model receptor binding site, the complex minimized, and binding energies calculated. These can then be compared with experimental binding affinities to see how well the model agrees with experiment. If the results make sense, the model can then be used for the design and synthesis of new agents.

The procedures described above are an example of a process known as 3D QSAR, which is described in section 18.10. Further details on specific software programs that generate hypothetical pseudoreceptors is given in section 18.10.5.

KEY POINTS

- The active conformation and pharmacophore of a molecule can be identified from the X-ray structure of a complex between it and its target protein. Alternatively, the active conformation and pharmacophore can be identified from an active, rigid structure. If no such rigid structures are available, the various conformations of different active compounds can be compared to identify areas of space and binding groups that are shared by all the active structures.
- Docking involves the fitting of a molecule into a binding site. Docking can be carried out on the basis of steric complementarity and/or chemical complementarity.
- Docking is most easily carried out with a rigid ligand and a rigid binding site. It is also possible to carry out docking with a flexible ligand and a rigid binding site.
- The docking of flexible ligands involves programs that use a fragmentation approach followed by the docking of a rigid anchor and reconstruction of the ligand. Alternatively, simulated annealing or genetic/evolutionary algorithms can be used to generate different conformations during the docking process.
- Electronic screening or database mining involves the search of structural databases to identify structures containing a particular pharmacophore.
- The binding sites of a target protein can be constructed by molecular modelling based on the X-ray structure of a protein–ligand complex. Alternatively, a model binding site can be constructed based on the primary sequence of the protein and a structural comparison of known analogous proteins. Another method is to compare a range of active compounds to identify where particular amino acids are likely to be, in order to allow an interaction.

17.15 *De novo* design

17.15.1 General principles of *de novo* design

De novo design involves the design of novel structures based on the structure of the binding site with which they are meant to interact. The structure of the binding site can be identified from an X-ray crystallographic study of the target protein containing a bound ligand or inhibitor. The position of the ligand identifies where the binding site is in the protein and also identifies any induced fit that might have occurred as a result of it binding. Once the structure of the protein–ligand complex has been downloaded onto a computer, the ligand can be removed to leave the empty binding site, and *de novo* design can then take place. By identifying the amino acids that are present in the binding site, it is possible to identify the binding interactions that are possible within the site. A drug can then be designed which will have the correct size and shape to fit the space available, and will also have the required functional groups to interact with the binding regions. The operator can carry out each of these operations manually. Alternatively, there are many software packages available which will carry out the process automatically.

Manual studies allow operators to have full control over the study and to input their own ideas as and when they wish. Such studies have been successful in producing novel active compounds, but they do suffer several disadvantages. For example, the novelty of the structures obtained is inevitably limited to the operator's own imagination and originality. More seriously, manual design is slow, and it is really limited to the identification of a single novel structure. Automatic design is much faster and can produce large numbers of diverse structures in a short time period.

The early work on *de novo* drug design was carried out manually, and an example of one of these studies is given in Case Study 5. From these studies, a number of general principles was identified regarding manual *de novo* design.

First, it may be tempting to design a molecule which completely fills the space that is available in the binding site. However, this would not be a good idea for the following reasons:

- Normally, the structure of the binding site is identified from X-ray crystallography of the target protein. The position of atoms in the crystal structure is accurate only to 0.2–0.4 Å, and allowance should be made for that.
- It is possible that the designed molecule may not bind to the binding site exactly as predicted. If the intended fit is too tight, a slight alteration in the binding mode may prevent the molecule binding at all. It would be better to have a loose-fitting structure in the first instance and to check whether it binds as intended. If it does not, the loose fit gives the molecule a chance to bind in an alternative fashion.
- It is worth leaving scope for variation and elaboration in the molecule. This allows fine-tuning of the molecule's binding affinity and pharmacokinetics.

Other important points to take into consideration in *de novo* design are the following:

- Flexible molecules are better than rigid molecules, because the former are more likely to find an alternative binding conformation should they fail to bind as expected. This allows modifications to be carried out based on the actual binding mode. If a rigid molecule fails to bind as predicted, it may not bind at all.
- It is pointless designing molecules that are difficult or impossible to synthesize.
- Similarly, it is pointless designing molecules that need to adopt an unstable conformation in order to bind.
- Consideration of the energy losses involved in water desolvation should be taken into account.
- There may be subtle differences in structure between receptors and enzymes from different species. This is significant if the structure of the binding site used for *de novo* design is based on a protein that is not human in origin.

These principles also hold true for automated *de novo* design and some are particularly problematic. For example, automated *de novo* design is prone to generating structures which are either difficult or impossible to synthesize. Consequently, efforts have been made to improve the software packages involved, such that they can identify and filter out problem structures, or prevent them being generated in the first place. A second problem with automated *de novo* programs revolves around the scoring functions used to estimate binding affinities. It would be useful to rank the generated structures with respect to their binding strengths, but the results obtained have often been found to be unreliable.

Critics of *de novo* drug design are quick to point out that no clinically useful drugs have been designed in this manner. This is true, but it is hard to see how this could be a realistic expectation. The number and variety of structures which could be identified through *de novo* design is virtually limitless and so the chances of 'hitting' the ideal structure are poor. Moreover, there is far more to drug design than finding a structure that binds strongly to its target. *De novo* drug design does not identify whether the structures identified will have favourable pharmacokinetic properties or acceptable safety profiles.

The real strength in *de novo* drug design is that it can stimulate new ideas and identify novel lead structures which could then be optimised through structure-based drug design (see for example section 20.7.4.4).

17.15.2 Automated *de novo* design

Several computer software programs have been written which automatically design novel structures to fit known binding sites. The following are some examples.

17.15.2.1 LUDI

One of the best known *de novo* software programs is called **LUDI,** which works by fitting molecular fragments to different regions of the binding site then linking the fragments together (Fig. 17.49). There are three stages to the process.

Stage 1: Identification of interaction sites

First of all, the atoms present in the binding site are analysed to identify those that can take part in hydrogen bonding interactions, and those that can take part in van der Waals interactions. Oxygen atoms and tertiary nitrogen atoms are identified as hydrogen bond acceptors. A hydrogen attached to oxygen or nitrogen is identified as a hydrogen bond donor. Aromatic and aliphatic carbons are identified as such, and are capable of taking part in van der Waals interactions.

Interaction sites can then be defined. These are positions in the binding site where a ligand atom could be placed to interact with any of the above atoms. For example, suppose the binding site contains a methyl group (Fig. 17.50). The program would identify the carbon of that group as an aliphatic carbon capable of taking part in van der Waals interactions. This is a non-directional interaction, so a sphere is constructed around the carbon atom with a radius corresponding to the ideal distance for such an interaction (4 Å). A number of points (normally 14) are then placed evenly over the surface of the sphere to define aliphatic interaction sites. Regions of the sphere which overlap or come too close to the atoms making up the binding site (i.e. less than 3 Å separation) are rejected, along with any of the 14 points that were on that part of the surface. The remaining points are then used as the aliphatic interaction sites. A similar process is involved in identifying aromatic interaction sites surrounding an aromatic carbon atom.

Identifying interaction sites for hydrogen bonds is carried out in a different fashion. As hydrogen bonds are directional, it is important to define not only the distance between the ligand and the binding region, but the relevant orientation of the atoms as well. This can be done by defining the hydrogen bond interaction site as a vector involving two atoms. The position of these atoms is determined by the ideal bond lengths and bond angles for a hydrogen bond. For example, if the binding site has a carbonyl group present, then there are two possible hydrogen bond interaction sites (X–H) that can be determined (Fig. 17.51).

If the binding site has a hydrogen bond donor present, then interaction sites for a hydrogen bond acceptor would be determined in a similar fashion. For example, Fig. 17.52 shows how the interaction site for a hydrogen

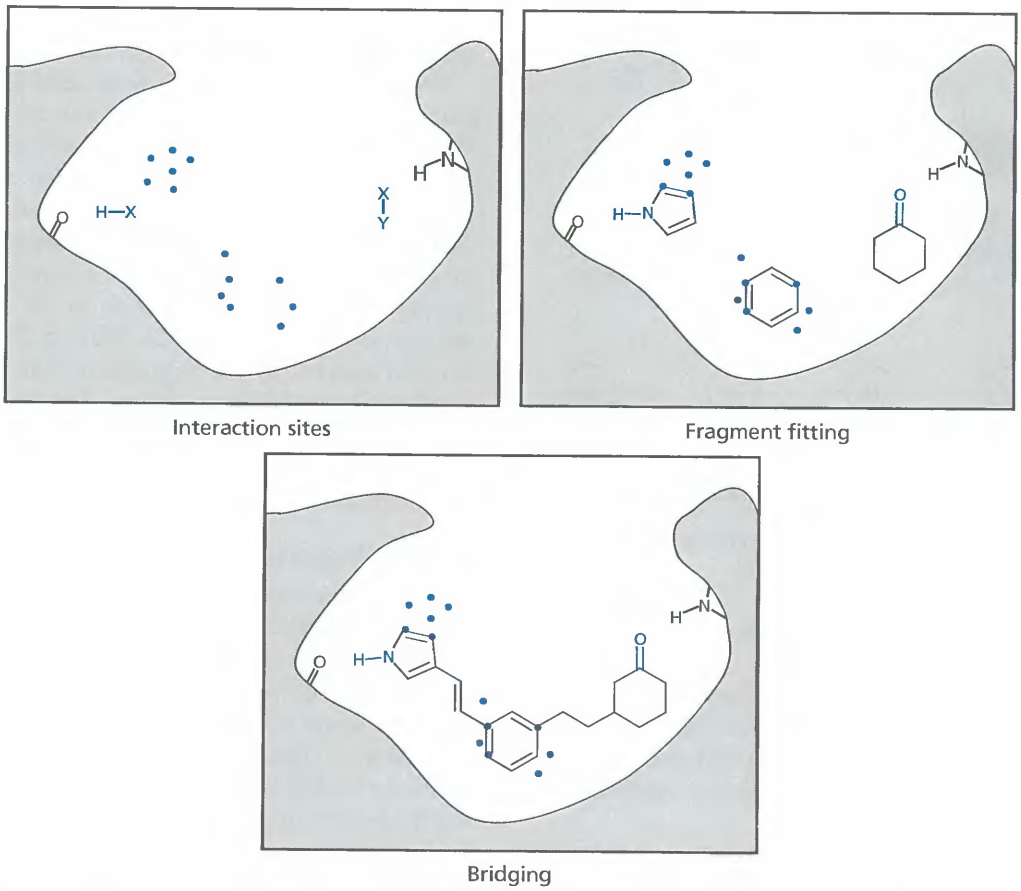


FIGURE 17.49 Stages involved in automated *de novo* drug design using LUDI. H–X, hydrogen bond donor interaction site; X–Y, hydrogen bond acceptor interaction site. The dots indicate aromatic interaction sites.

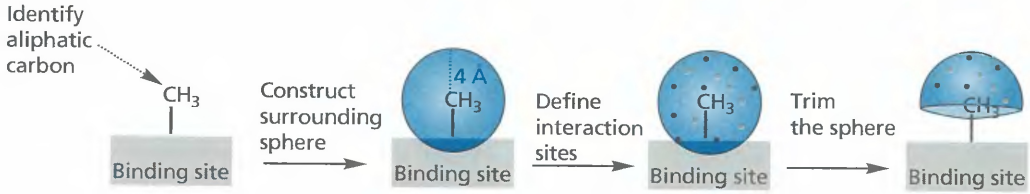


FIGURE 17.50 Identification of aliphatic interaction sites round a methyl group (LUDI).

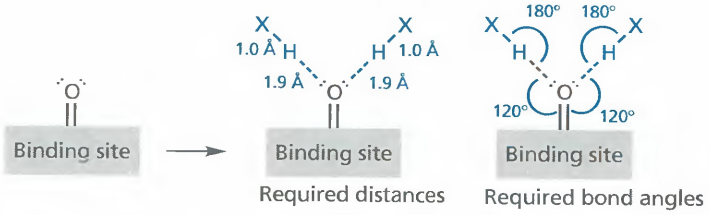


FIGURE 17.51 The interaction sites for a hydrogen bond donor, represented by H–X (LUDI).

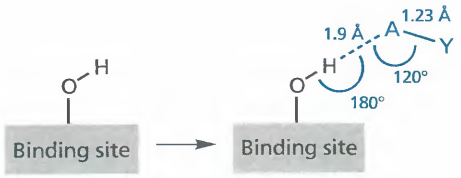


FIGURE 17.52 The interaction site for a hydrogen bond acceptor represented by A–Y (LUDI). A denotes the hydrogen bond acceptor.

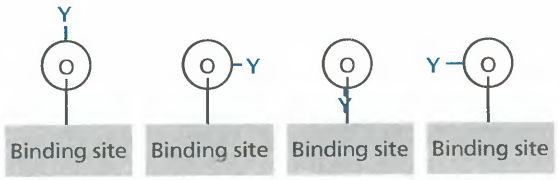


FIGURE 17.53 Four possible interaction sites for A–Y (A is hidden).

bond acceptor is determined when the binding site has a hydroxyl group present.

There are in fact four interaction sites that are normally calculated in this situation. The other three can be visualized if we take a viewpoint along the line of atoms O–H–A and vary the relative position of Y as shown (Fig. 17.53).

As with the van der Waals interaction sites, hydrogen bond interaction sites are checked to ensure that they are no closer than 1.5 Å from any atom present in the binding site. If they are, they are rejected.

Stage 2: Fitting molecular fragments

Once interaction sites have been determined, the LUDI program accesses a library of several hundred molecular fragments such as those shown in Fig. 17.54. The molecules chosen are typically 5–30 atoms in size and are usually rigid in structure because the fitting procedure assumes rigid fragments. Some fragments are included which *can* adopt different conformations. For these fragments, a selection of different conformations has to be present in the library if they are to be represented fairly in the fitting process. Each conformation is treated as a separate entity during the fitting process.

The atoms which are going to be used in the fitting process have to be predetermined for each fragment. Similarly, the interaction sites on to which each atom can be fitted have to be predetermined. For example, the methyl carbons of an acetone fragment are defined as aliphatic and can only be fitted onto aliphatic interaction sites. The carbonyl group is defined as a hydrogen bond acceptor and can only be fitted on to the corresponding interaction site (Fig. 17.55). The best fit will be the one that matches up the fragment with the maximum number of interaction sites. The program can 'try out' the various fragments in its library and identify those that can be matched up or fitted to the available interaction sites in the binding site.

Stage 3: Fragment bridging

Once fragments have been identified and fitted to the binding site, the final stage is to link them up. The program first identifies the molecular fragments that are closest to each other in the binding site, then identifies the closest hydrogen atoms (Fig 17.56). These now define the link sites for the bridge. The program now tries out various molecular bridges from a stored library to find out which one fits best. Examples of the types of molecular bridges that are stored are shown in Fig. 17.57. Once a suitable bridge has been found, a final molecule is created.

17.15.2.2 SPROUT

Another early example of an automated **de novo** drug design program is **SPROUT**. Like LUDI, the programme fits fragments to interaction sites, but there are interesting differences in the way that the process is carried out. For example, the interaction sites that are used in the programme consist of atom-sized spheres. The spheres represent a volume of space within the binding site into which a ligand atom should be placed in order to interact favourably through hydrogen bonding or van der Waals interactions. Alternatively, spheres can be placed into the binding

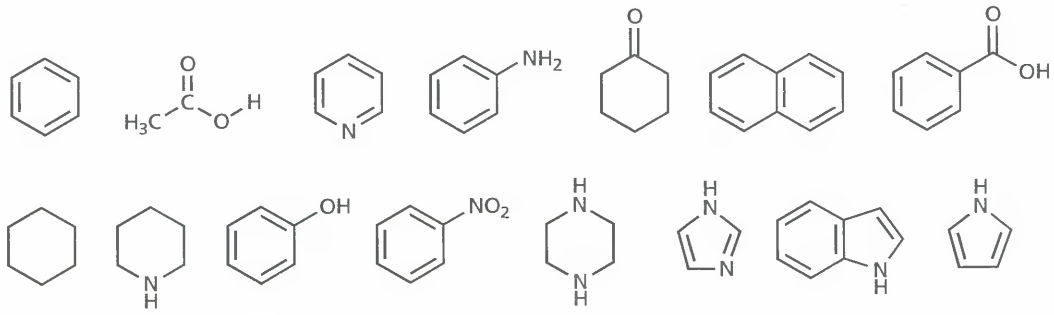


FIGURE 17.54 Examples of molecular fragments used by LUDI.

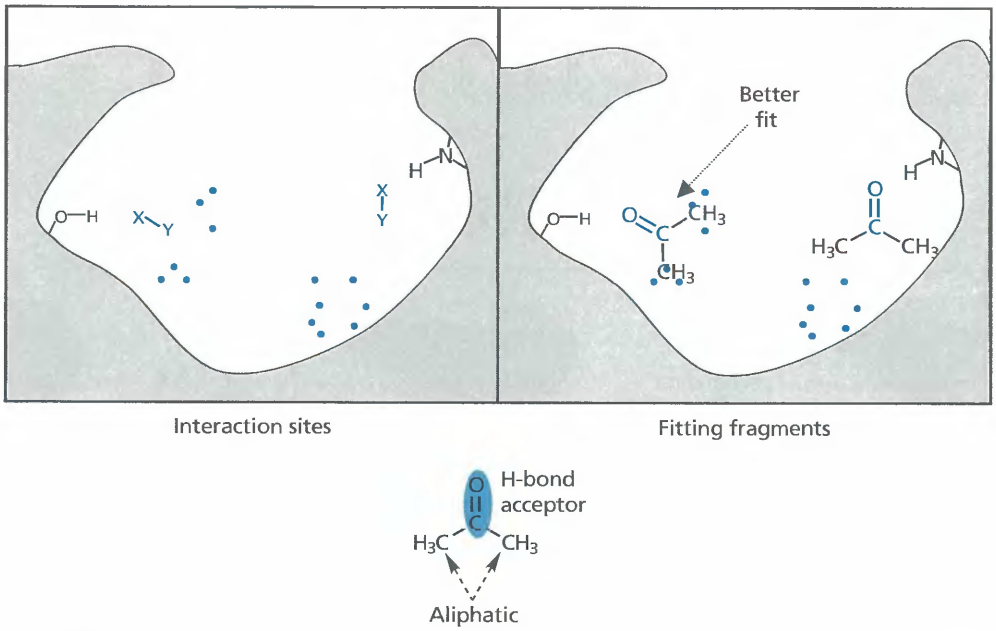


FIGURE 17.55 Fitting fragments. X–Y, hydrogen bond acceptor interaction site. The dots indicate aliphatic interaction sites.

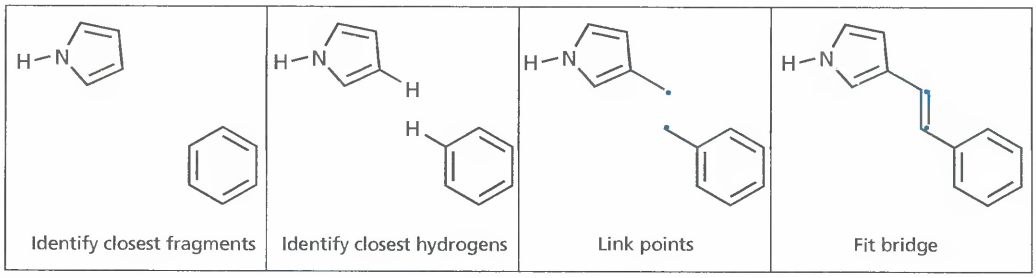


FIGURE 17.56 The bridging process (LUDI).

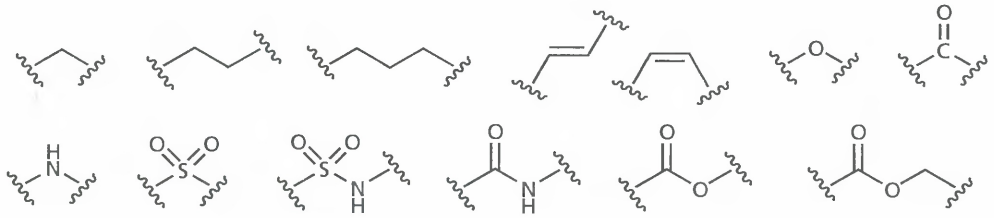


FIGURE 17.57 Examples of molecular bridges (LUDI).

site to ensure that a particular structural feature is present in the final structures—for example an aromatic ring.

As far as the 'building blocks' are concerned, SPROUT uses fragment templates to represent molecular fragments. We have already seen examples of molecular fragments used in LUDI. The atoms and bonds are specified in these fragments and there are a huge number of possible fragments which could be considered. Molecular templates on the other hand are designed to represent several different fragments. Each molecular template is defined by vertices and edges rather than by atoms and bonds. A vertex represents a generalised sp, sp^2 or sp^3 hybridized atom, while an edge represents a single, double or triple bond, depending on the hybridization of the vertices at either end. For example, the fragment template shown in Fig. 17.58 can represent a large number of different 6 membered rings. This approach has the advantage that it radically cuts down the number of different fragments that have to be stored in the program, making the search for novel structures more efficient. However, there is no

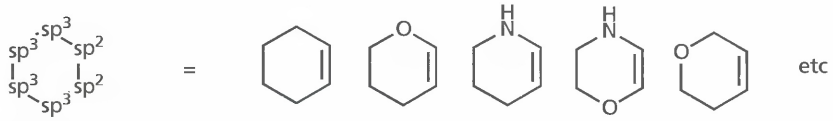


FIGURE 17.58 Examples of structures represented by a molecular template used in SPROUT.

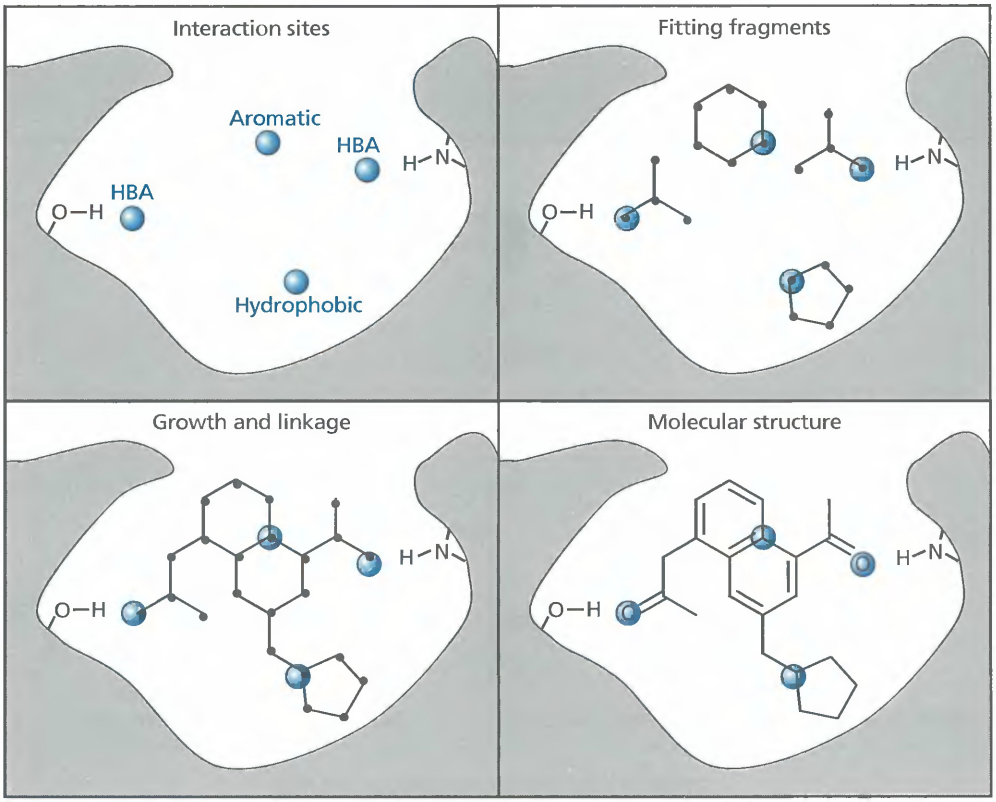


FIGURE 17.59 Generating structures using SPROUT (HBA = hydrogen bond acceptor).

reason why specific templates cannot be used as well, and the current version of SPROUT allows a mixture of specific molecular fragments and generic templates to be used at the same time.

The generation of the structures takes place in two stages (Fig. 17.59). In the first stage, the emphasis is on generating molecular templates that will fit the binding site. There is no consideration of binding interactions at this stage, and so there is no need to know what sort of atoms are present in the fragment templates. The program selects a fragment template randomly and positions it into the binding site by placing one of the vertices at the centre of a sphere. In the early versions of SPROUT, further fragments were then added sequentially, and the skeleton was 'grown' until it occupied all the other spheres. In the current version, fragments are placed at all the spheres and are grown towards each other, until they are finally linked. One advantage of SPROUT is that the 'growth' of fragment templates allows a molecular template to be constructed which bridges interaction sites that are some distance apart. In the LUDI method, single templates are placed at each interaction point and are then linked. If there is a large separation between the interaction sites, there might not be a sufficiently long linker to connect the fragments.

The second stage in the process is to create molecules from the template skeletons that have been produced. This involves replacing the vertices with suitable atoms to allow favourable hydrogen bonding and van der Waals interactions with the binding site. For example, if a vertex is located within a sphere that requires a hydrogen bond acceptor, an oxygen or a nitrogen atom can be added at that position. Since generic templates have been used to generate each skeleton, a large variety of molecular structures can be generated from each molecular template.

SPROUT has the capacity to identify certain structural features that might be unrealistic and then modify them. For example, an OH might be generated during the second stage in order to introduce a hydrogen bond donor, but if the OH is linked to a double bond this results in an enol which would tautomerise to a ketone. The latter would not be able to act as a hydrogen bond donor. The programme can identify an enol and modify it to a carboxylic acid which can still act as a hydrogen bond donor (Fig. 17.60).

The programme also has the ability to modify structures such that they are more readily synthesized. For example, introducing a heteroatom into a two-carbon link between two rings generates a structure which can be more readily synthesized (Fig. 17.61). In this example, the link could be made synthetically by reacting an alkoxide with an alkyl halide (Fig. 17.62).

The structures that are finally generated by SPROUT are then evaluated *in silico* for a variety of properties including possible toxicity and pharmacokinetic properties.

The program **CAESA** is used to evaluate how easily each structure can be synthesized and to give an indication of likely starting materials for the synthesis. The program does this by carrying out a retrosynthetic analysis of each structure.

More recently, SPROUT has introduced a method of assessing the synthetic feasibility of the partial structures created during the *de novo* construction process. Such an analysis is useful since it allows a pruning process to take place which rejects partial structures that are not easily synthesized, and directs the program to generate more suitable structures. CAESA itself cannot be used for this purpose since it is relatively slow, taking about a minute per structure. This is acceptable for the analysis of the final structures that are generated, but would slow up the process considerably if it was used to assess the thousands of intermediate structures that are generated during the building process. Therefore, a less accurate but quicker method of analysis is carried out. The method involves the identification of molecular

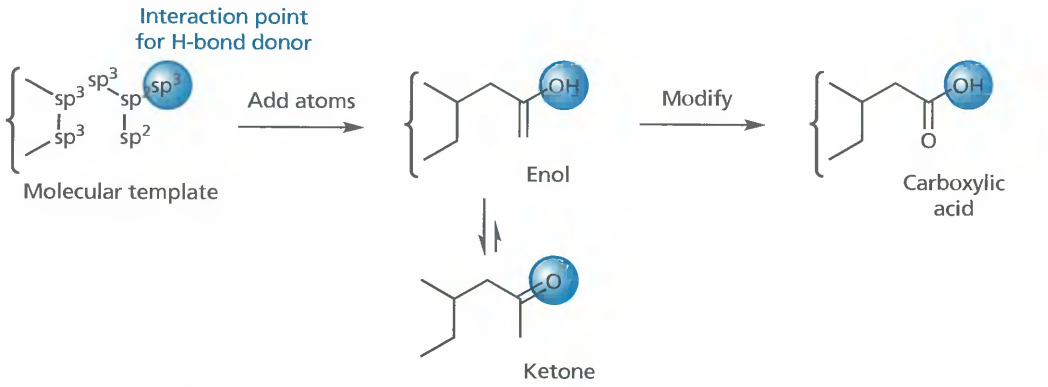


FIGURE 17.60 Modification of an enol to a carboxylic acid by SPROUT.

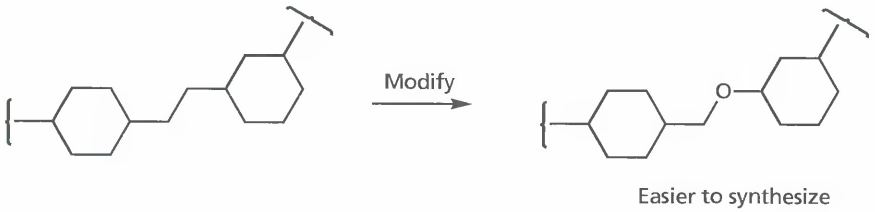


FIGURE 17.61 Modification by SPROUT to generate a more synthetically feasible structure.

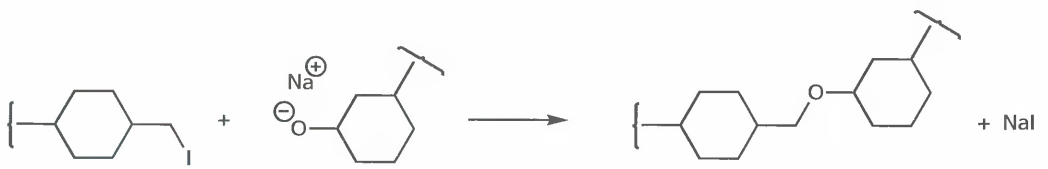


FIGURE 17.62 Possible synthesis allowing the linkage required in Fig. 17.61.

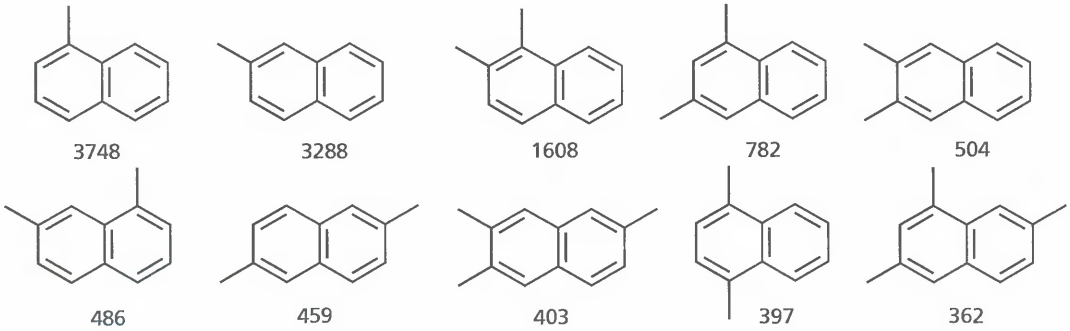


FIGURE 17.63 The ten most frequent substitution patterns for naphthalene in a database of drug-like compounds. Numbers refer to number of occurrences.

features within each partial structure and identifying how frequently they occur in known structures. The rationale is that if a particular feature commonly exists in known compounds, it is likely that that same feature should be capable of synthesis in the novel structures generated by *de novo* design.

The major structural features within a molecule can be defined as the various sized rings that are present, as well as any connecting chains. The synthetic feasibility of rings and chains is generally dependent on their substitution pattern. For example, the ten most frequent substitution patterns for a naphthalene ring amongst a database of active compounds are shown in Fig. 17.63. The analysis of partial structures can be carried out such that structures with uncommon structural features (such as a tetrasubstituted naphthalene ring) are penalised and rejected. A measure of how drug-like the partial structures are can also be gleaned if the database used in the analysis is restricted to active compounds from drug databases.

17.15.2.3 LEGEND

LEGEND is another long established automated *de novo* design program. A grid is set up within the binding site to identify steric and electrostatic interaction energies between each grid point and the binding site (section 17.7.5). These are tabulated for different types of atom and are used to estimate van der Waals interactions for the growing skeletons that are generated by the program, as well as for structure optimization of final structures. The operator has the choice of starting from a single heteroatom, placed in such a position that it can form a hydrogen bond with the binding site. Alternatively, a molecule or molecular fragment can be placed into the binding site to act as a starting structure. This can be useful if one wants to include the partial structure of an active compound within the generated structures. Once the starting atom or fragment has been positioned, the growth stage can commence to generate different skeletons.

Unlike LUDI and SPROUT, LEGEND does not use fragments or templates to generate skeletons. Instead, the skeletons are grown one atom at a time using random choices at each stage of the process. For example, the type of atom to be added is chosen at random. The root atom (the atom in the existing skeleton to which the new atom is to be linked) is also chosen at random, as is the type of bond used for the connection. Finally, a random torsion angle is chosen to position the new atom relative to the existing skeleton. Particular features such as aromatic rings, carbonyl groups and amide groups can be generated since some of the atom types that are used are defined as belonging to these features. For example, if a new atom is defined as being an aromatic carbon, then the final structure must eventually contain an aromatic ring containing that atom. The aromatic ring may not be completely formed when the growth stage is over, but the program will automatically complete the ring.

This approach of adding atoms one by one has the advantage that it can generate a greater diversity of structures than those generated by fragment-based procedures. However, it suffers from the disadvantage that the number of different structures generated can increase dramatically as each atom is joined. For that reason, it is important to evaluate the growing structures at each stage of the process, and to carry out pruning operations. This inevitably means that the generation of structures is a slower process compared to fragment-based methods. As each atom is added, the structure is checked to ensure that there are no steric clashes within the molecule itself or between the molecule and the binding site. This is done by measuring the van der Waals interactions using the tabulated values obtained from the grid measurements. If these *are* unfavourable, the structure is rejected and the program backtracks to choose a different root atom. If that also fails to generate an acceptable skeleton, the last atom to be added to the skeleton is removed and a new root atom is chosen.

When a new skeleton is accepted, the position of the new atom is assessed to see if it lies on a grid point

associated with a very large electrostatic potential. If this is the case, the program ensures that it is changed to a heteroatom such that a hydrogen bond or ionic bond is possible with the binding site. The skeleton continues to grow until it reaches a size that is pre-determined by the operator. At that stage, hydrogen atoms are added to complete the valencies of each atom. If partially constructed aromatic rings are present, these are also completed. The structure is finally optimised taking into account both intramolecular and intermolecular interactions. The process is then repeated to generate as many structures as desired.

17.15.2.4 GROW and SYNOPSIS

GROW is a program that uses molecular fragments to generate novel ligands for binding sites. The fragments used represent amino acids, and so the structures that are generated are limited to peptides.

SYNOPSIS is a *de novo* design program that is designed to generate synthetically feasible structures. It does so by incorporating synthetic rules into the structure building process. In other words, fragments can only be linked if there is a known reaction which will allow it. Moreover, the fragments used must be commercially available. This program not only generates synthetically feasible molecules, but also provides a possible synthetic route.

KEY POINTS

- *De novo* design involves the design of a novel ligand, based on the structure of the binding site.
- An X-ray crystal structure of the target protein complexed with a ligand allows identification of the binding site and the binding mode of the ligand.
- A new ligand should initially be loose fitting and flexible to allow for any alterations in the way binding takes place compared to what is predicted.
- The X-ray structure of the protein complexed with the new ligand will give valuable information on the actual binding mode of the new ligand and allow modifications to be made which will maximize bonding.
- The new ligand should be capable of synthesis and be capable of interacting with the binding site using a stable conformation.
- Energy losses resulting from desolvation of the ligand should be taken into account when calculating stabilization energies arising from ligand–protein binding.
- Automated programs for *de novo* design identify interaction sites in the binding site, then match molecular fragments to these sites. Bridges are then designed to link the fragments.

17.16 Planning combinatorial synthesis

Combinatorial synthesis (chapter 16) is a method of creating a large number of compounds on small scale and in a short time. The combinatorial synthesis could be carried out to synthesize all the possible compounds from the available starting materials and reagents. However, molecular modelling can help to focus the study so that a smaller number of structures are made but the chances of success are retained.

One method of doing this is based on the identification of **pharmacophore triangles** (section 17.11.3). Let us assume that the combinatorial synthesis is being carried out to generate 1000 compounds with as diverse a range of structures as possible. The number of different pharmacophores generated from the 1000 compounds would be an indication of the structural diversity. Therefore, a library of compounds which generates 100 000 different pharmacophores would be superior to a library of similar size which produces only 100 different pharmacophores. Pharmacophore searching can be done on all the possible synthetic targets to select those structures that will give the widest diversity. These compounds would then be synthesized.

First, all the possible synthetic targets are automatically ranked on their level of rigidity. This can be achieved by identifying the number of rotatable bonds. Pharmacophore searching then starts with the most rigid structure. All the possible pharmacophore triangles are identified for the structure. If different conformations are possible, these are generated and the various pharmacophore triangles arising from these are added to the total. The next structure is then analysed for all of its pharmacophore triangles. Again, triangles are identified for all the possible conformations. The pharmacophores from the first and second structures are then compared. If more than 10% of the pharmacophores from the second structure are different from those of the first, both structures are added to the list for the intended library. Both sets of pharmacophores are combined, and the next structure is analysed for all of its pharmacophores. These are compared with the total number of pharmacophores from structures 1 and 2 and if there are 10% new pharmacophores represented, the third structure is added to the list and the pharmacophores for all three structures are added together for comparison with the next structure. This process is repeated throughout all the target structures, eliminating all compounds which generate less than 10% of new pharmacophores. In this way, it is possible to cut the number of structures which need to

be synthesized by 80–90% with only a 10% drop in the number of pharmacophores generated.

There is a good reason for starting this analysis with a rigid structure. A rigid structure has only a few conformations and there is a good chance that most of these will be represented when the structure interacts with its target. Therefore, one can be confident that the associated pharmacophores are also represented. If the analysis started with a highly flexible molecule having a large number of conformations, there is less chance that all the conformations and their associated pharmacophores will be fairly represented when the structure meets its target binding site. Rigid structures which express some of these conformations more clearly would not be included in the library, as they would be rejected during the analysis. As a result, some pharmacophores which should be present are actually left untested.

It is possible to use modelling software to carry out a substituent search when planning a combinatorial synthetic library. Here, one defines the common scaffold created in the synthesis, as well as the number of substituents which are attached and their point of attachment. Next, the general structures of the starting materials used to introduce these substituents are defined. The substituents which can be added to the structure can then be identified by having the computer search databases for commercially available starting materials. The program then generates all the possible structures which can be included in the library, based on the available starting materials. Once these have been identified, they can be analysed for pharmacophore diversity as described above.

Alternatively, the various possible substituents can be clustered into similar groups on the basis of their structural similarity. This allows starting materials to be pre-selected choosing a representative compound from each group. The structural similarity of different substituents would be based on a number of criteria such as the distance between important binding centres, the types of centre present, particular bonding patterns, and functional groups.

17.17 Database handling

The development of a drug requires the analysis of a large amount of data. For example, activity against a range of targets has to be measured to ensure that the compounds have not only good activity against their intended target, but also good selectivity with respect to a range of other targets. When it comes to rationalizing results, many other parameters have to be considered such as molecular weight, log P, and pK_a. The handling of such large amounts of data requires dedicated software.

Several software programs are available for the handling of data which allow medicinal chemists to assess biological activity versus physical properties, or to compare the activities of a series of compounds at two different targets. Such programs permit results to be presented in a visual qualitative fashion, allowing a quick identification of any likely correlations between different sets of data.

For example, if one wanted to see whether the log P value of a series of compounds was related to their α- or β-adrenergic activity, then a 2D plot could be drawn up comparing α-adrenergic activity versus β-adrenergic activity. The log P value of each compound could then be indicated by a colour code for the various points on the plot. In this way, it would be easy to see whether these three properties were related. Such an analysis might show for example that a high log P is associated with compounds having low α-adrenergic activity and high β-adrenergic activity.

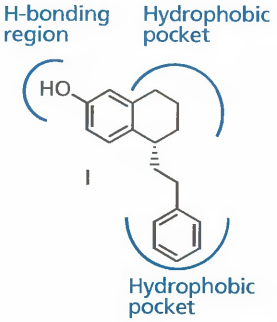
Some programs can be used to assess the biological results from a combinatorial synthetic study. First, the scaffold used in the synthesis is defined, then the substituents are defined. Once the biological test results are obtained, a tree diagram can be drawn up to assess which substitution point is most important for activity. For example, supposing there were three substitution points on the scaffold, the program could analyse the data to identify which of the substitution points was the most important in controlling the activity. The data relevant for this particular substituent could then be split into three groups corresponding to good, average and poor activity. For each of these groups, the program could be used to identify the next most important substitution point, and so on.

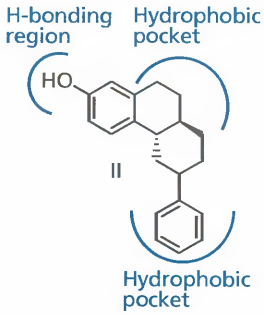
KEY POINTS

- Molecular modelling can be used to plan intended combinatorial libraries such that the maximum number of pharmacophores are generated for the minimum number of structures.
- Structures are analysed for their various conformations and resulting pharmacophores, starting with the most rigid structures.
- Each structure is compared with a growing bank of pharmacophores to assess whether it presents a significantly different number of pharmacophores compared to the structures that went before.

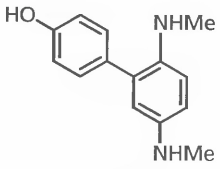
QUESTIONS

1. What is meant by energy minimization and how is it carried out?
2. What is meant by the terms local and global energy minima, and what relevance have they to conformational analysis?
3. What two properties should be known about two drugs if they are to be overlaid as a comparison?
4. Is it reasonable to assume that the most stable conformation of a drug is the active conformation?
5. You are carrying out a *de novo* design of a ligand for a binding site that contains a hydrogen bonding region and two hydrophobic pockets. Structures I and II below are both suitable candidates. Compare the relevant merits of these structures and decide which one you would synthesize first to test your binding theory.

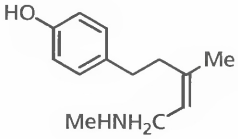




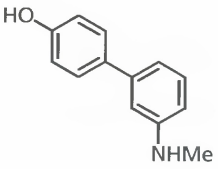
6. Both structures I and II above show poor water solubility. It is suggested that the phenyl group be replaced by a pyridine ring? What would be the advantages and disadvantages of this idea? Have you any alternative ideas?
7. Assuming that structures I and II both bound to the binding site as predicted, what further modifications might you make to increase binding interactions?
8. Why were such modifications not carried out earlier?
9. The following eight structures have been tested for activity as receptor agonists. Five are active and three are inactive. Assess the structures and discuss what the pharmacophore might be for agonist activity.



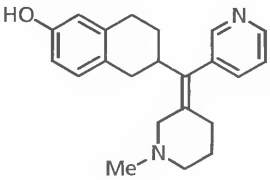
Active



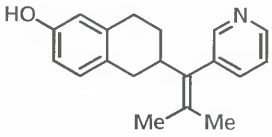
Active



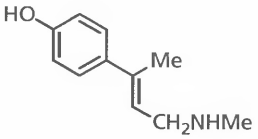
Active



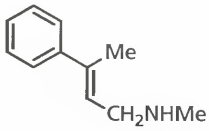
Active



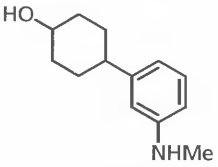
Inactive



Active



Inactive



Inactive

10. How would you go about carrying out overlays of the active structures in Question 9?

FURTHER READING

Agrafiotis, D. K., Lobanov, V. S., and Salemme, F. R. (2002) Combinatorial informatics in the post-genomics era. *Nature Reviews Drug Discovery*, **1**, 337–346.

Bikker, J. A., Trumpp-Kallmeyer, S., and Humblet, C. (1998) G-protein coupled receptors: models, mutagenesis, and drug design. *Journal of Medicinal Chemistry*, **41**, 2911–2927.

Boda, K., and Johnson, A. P. (2006) Molecular complexity analysis of de novo designed ligands. *Journal of Medicinal Chemistry*, **49**, 5869–5879.

Bohm, H.-J. (1992) The computer program LUDI: A new method for the *de novo* design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design*, **6**, 61–78.

Bourne, P. E., and Wessig, H. (eds.) (2003) *Structural bioinformatics*. John Wiley and Sons, New York.

Brooijmans, N., and Kuntz, I. D. (2003) Molecular recognition and docking algorithms. *Annual Rev. Biophys. Biomol. Struct.*, **32**, 335–73.

Ganellin, C. R., and Roberts, S. M. (eds.) (1994) Discovery and development of cromokalim and related potassium channel activators. Chapter 8 in *Medicinal chemistry—the role of organic research in drug research*, 2nd edn. Academic Press, New York.

Greer, J., Erickson, J. W., Baldwin, J. J., and Varney, M. D. (1994) Application of the three-dimensional structures of protein target molecules in structure-based design. *Journal of Medicinal Chemistry*, **37**, 1035–1054.

Kitchen, D. B., *et al.*, Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, **3**, 935–949.

Kobilka, B., and Schertler, G. F. X. (2008??), New G-protein-coupled receptor crystal structures: insights and limitations. *Trends in Pharmacological Sciences*, **29**, 79–83.

Leach, A. R. (2001) *Molecular modelling: principles and applications*, 2nd edn. Pearson Education, London.

Miller, M. A. (2002) Chemical database techniques. *Nature Reviews Drug Discovery*, **1**, 220–227.

Navia, M. A., and Murcko, M. A. (1992) Use of structural information in drug design. *Current Opinion in Structural Biology*, **2**, 202–216.

Richards, G. (2002) Virtual screening using grid computing: the screensaver project. *Nature Reviews Drug Discovery*, **1**, 551–555.

Schlyer, S., and Horuk, R. (2006) I want a new drug: G-protein-coupled receptors in drug development. *Drug Discovery Today*, **11**, 481–493.

Schneider, G., and Fechner, U. (2005). Computer-based *de novo* design of drug-like molecules. *Nature Reviews Drug Discovery*, **4**, 649–663.

Shoichet, B. K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862–865.

van de Waterbeemd, H., Testa, B., and Folkers, G. (eds.) (1997) *Computer-assisted lead finding and optimization*. Wiley-VCH, New York.

Titles for general further reading are listed on p. 725.